THE ACCOUNTING REVIEW

VOLUME 86, No. 1 JANUARY 2011

A JOURNAL OF THE
AMERICAN ACCOUNTING ASSOCIATION

AMERICAN ACCOUNTING ASSOCIATION

The American Accounting Association was founded in 1916 as the American Association of University Instructors in Accounting. Its present name was adopted in 1936. The Association is a voluntary organization of persons interested in accounting education, research, and practice.

Membership in the Association is available at the following annual rates:

Full membership:
The membership dues include *Accounting Education News*, and online selection of one, two, or all three of the following Association-wide journals: *The Accounting Review*, *Accounting Horizons*, and *Issues in Accounting Education*. (This does not include Section journals.) Select from:

The Accounting Review only	U.S. $185.00	with Print $220.00
Accounting Horizons OR *Issues in Accounting Education* only	U.S. $185.00	with Print $200.00
The Accounting Review AND *Accounting Horizons* OR *Issues in Accounting Education*	U.S. $195.00	with Print $245.00
Accounting Horizons AND *Issues in Accounting Education*	U.S. $195.00	with Print $225.00
All three Association-wide journals	U.S. $205.00	with Print $270.00

Associate membership:
Full-time students are eligible for associate membership. The print option is not available for Associate members. Select from:

Any one Association-wide journal	U.S. $25.00
Any two Association-wide journals	U.S. $35.00
All three Association-wide journals	U.S. $45.00

In addition, full members may belong to one or more Sections: Accounting, Behavior and Organizations; Accounting Programs Leadership Group; American Taxation Association; Auditing; Diversity; Financial Accounting and Reporting; Forensic and Investigative Accounting; Gender Issues and Worklife Balance; Government and Nonprofit; Information Systems; International Accounting; Management Accounting; Strategic and Emerging Technologies; Teaching, Learning and Curriculum; and Two-Year College. Each Section assesses its own annual dues, which are collected by the Association.

Editorial correspondence and all manuscripts should be emailed to tar@mccombs.utexas.edu. If it is necessary to send any hard copy correspondence, the mailing address is Mary Capps, Editorial Assistant, *The Accounting Review*, The University of Texas at Austin; Department of Accounting, CBA 4M202; 1 University Station B6400; Austin, TX 78712-0211 • Telephone: (512) 471-5318 • Fax: (512) 471-3904.

The submission fee of $200.00 for members of the AAA or $400.00 for nonmembers should be paid online (VISA or MasterCard only) at: https://aaahq.org/AAAforms/journals/tarsubmit.cfm and must be paid when article is submitted. Request for reprint permission, subscriptions, advertisements, and dues should be sent to American Accounting Association, 5717 Bessie Drive, Sarasota, FL 34233-2399 • Telephone: (941) 921-7747 • Fax: (941) 923-4093 • Email: Info@aaahq.org. Notice of change of address should be mailed at least eight weeks in advance to ensure undelayed receipt of *The Accounting Review*.

The Accounting Review is indexed in the Social Science Citation Index which is included in Web of Science.

The Accounting Review (ISSN 0001-4826 print and ISSN 1558-7967 online) is published six times a year in January, March, May, July, September, and November by the American Accounting Association, 5717 Bessie Drive, Sarasota, Florida 34233-2399. General subscriptions are available at the annual rates of $395 for a print copy or at $445 for a print and electronic copy. The member subscription rate for a hard copy is $35. Electronic access is included with membership dues. Periodical postage paid at Sarasota, Florida and additional mailing offices. **POSTMASTER:** Send address change to *The Accounting Review*, 5717 Bessie Drive, Sarasota, Florida 34233-2399.

THE ACCOUNTING REVIEW

A JOURNAL OF THE

AMERICAN ACCOUNTING ASSOCIATION

VOL. 86 JANUARY 2011 No. 1

PAST EDITORS

1926–1929	William A. Paton
1929–1943	Eric L. Kohler
1944–1947	A. C. Littleton
1948–1949	Robert L. Dixon
1950–1959	Frank P. Smith
1959–1962	Robert K. Mautz
1962–1965	Lawrence L. Vance
1965–1967	Wendell Trumbull
1968–1970	Charles H. Griffin
1971–1972	Eldon S. Hendriksen
1973–1975	Thomas F. Keller
1976–1978	Don T. DeCoster
1978–1983	Stephen A. Zeff
1983–1987	Gary L. Sundem
1987–1990	William Kinney
1990–1994	A. Rashad Abdel-khalik
1994–1997	Robert P. Magee
1997–2000	Gerald L. Salamon
2000–2002	Linda Smith Bamber
2002–2005	Terry Shevlin
2005–2008	Dan S. Dhaliwal

THE ACCOUNTING REVIEW
Vol. 86, No. 1
2011
pp. 1–22

American Accounting Association
DOI: 10.2308/accr.00000015

2009 American Accounting Association
Competitive Manuscript Award Winner

Determinants of Investors' Information Acquisition: Credibility and Confirmation

Jane Thayer
The University of Georgia

ABSTRACT: Psychology theory suggests that current investors are unlikely to be impartial in their acquisition of information regarding a currently held position after receiving information that casts doubt on the profitability of the position. This study examines whether the favorability of initial information an investor receives subsequent to taking an investment position moderates his acquisition of additional information regarding that position. A web-based experiment indicates that participants generally choose to view information based on its credibility. However, the majority of information chosen by participants initially receiving unfavorable information supports the investment position they chose at the beginning of the case. Moreover, participants receiving unfavorable information prefer to view lower-credibility, preference-consistent information at a similar rate to higher-credibility, preference-inconsistent information. Given a psychological need to support a previous investment decision, participants forgo a certain amount of credibility in the information they gather to confirm their belief in their investment position.

Keywords: *investors; information acquisition; preferences; credibility; experiment.*

Data Availability: *Contact the author.*

I. INTRODUCTION

Investors have an economic incentive to be objective and impartial in their acquisition of information, as a high-quality set of information is central to an optimal investment decision. Psychology research suggests, however, that impartiality will be more difficult after an investment position has been taken (e.g., Festinger 1957; Leventhal and Brehm 1962; Festinger 1964).

This paper is based on my dissertation completed at Emory University. I thank my dissertation chairperson, Kathryn Kadous, for her support and guidance throughout this project, and my other dissertation committee members, Kristy Towry and Greg Waymire, for their many helpful suggestions. This paper also benefited from the valuable comments of Ben Ayers, Ann Backof, Linda Bamber, Michael Bamber, Brooke Elliott, Sukari Farrington, Jeffrey Hales, Jackie Hammersley, Gary Hecht, Max Hewitt, Pat Hopkins, Lisa Koonce, Susan Krische, Bob Libby, Laureen Maines, Molly Mercer, Mark Nelson, Derek Oler, Mark Peecher, Grace Pownall, Bill Tayler, Xue Wang, Donnie Young, and workshop participants at Carnegie Mellon University, Cornell University, Emory University, Georgia Institute of Technology, Indiana University, The University of Georgia, University of Illinois at Urbana–Champaign, University of Massachusetts Amherst, The University of Texas at Austin, and the 2008 American Accounting Association Annual Meeting. I am grateful for the financial support of the Roberto C. Goizueta Foundation and the Professor Jagdish and Mrs. Madhuri Sheth Fund.

Editor's note: Accepted by Laureen Maines.

Submitted: February 2009
Accepted: June 2010
Published Online: January 2011

An underlying assumption of cognitive dissonance theory is that individuals prefer their actions to match their beliefs (Festinger 1957). Information received subsequent to taking a position influences whether this internal balance is maintained (Festinger 1957, 1964). If information casts doubt on the future profitability of an investor's position, then psychology theory suggests that the investor may selectively seek additional information supporting the position to regain that balance (e.g., Frey 1986). This study examines whether the initial information an investor receives subsequent to taking an investment position influences the type of information s/he acquires to make future judgments and decisions about the position. Will an investor acquire a balanced set of credible information or will s/he acquire information that confirms his/her investment position?[1]

Investors' reactions to information based on its source credibility (e.g., Clement and Tse 2003; Gleason and Lee 2003) suggest that investors recognize the need to be accurate in their estimation of security prices. For example, investors react more quickly to forecast revisions made by celebrity analysts (Gleason and Lee 2003) and analysts employed by prestigious brokerage firms (Clement and Tse 2003) than to those made by lesser-known analysts.[2] In addition, investors react to the credibility of management forecasts in addition to the news, or surprise, in those forecasts (Jennings 1987; Mercer 2004; Hutton and Stocken 2009). Investors have also been shown to respond more strongly to earnings released by companies employing large, international audit firms compared to those of companies employing smaller audit firms (Teoh and Wong 1993). It is unknown, however, whether an investor's sensitivity to the credibility of information will continue after s/he receives initial information that casts doubt on a recently chosen position.

Although it is more difficult to remain objective after taking a position, research suggests that individuals are mindful of the credibility of information regarding the position, especially when the information is unfavorable (Ditto et al. 1998). However, given the choice between high-credibility information that is consistent with their preference to believe in the position (hereafter, preference-consistent information) and that which is inconsistent with their preferred belief (hereafter, preference-inconsistent information), studies show than an individual's need to support a desired belief will lead him/her to choose the former (Frey 1981, 1986).[3] These findings, however, result from a context that does not include an immediate economic incentive to gather a balanced set of credible information as an aid to decision making (Frey 1981).

With regard to individuals' willingness to trade off credibility for preference-consistency, studies suggest that individuals selectively seeking preference-consistent information will prefer to receive preference-inconsistent information if it has more credibility than available preference-consistent information (Lowin 1969). However, the psychology studies investigating individuals' preferences for information provide participants with information varied between two extreme levels of credibility (e.g., Lowin 1967, 1969; Frey 1981; Ditto et al. 1998). Given the vast amount of information from various sources in the capital markets, it is unknown whether an investor would seek information based on its credibility or whether s/he would choose to acquire information supportive of his/her investment position.

---

[1] Regarding information acquisition, I use the term "balanced" to describe a representative sample of the information available from all sources, with no one viewpoint under- or over-represented in the sample, relative to its representation in the population of available information.

[2] Studies show that rankings and employer size are related to analysts' forecast accuracy (Stickel 1992; Clement 1999; Hong and Kubik 2003); therefore, these signals can provide investors with a cognitive shortcut in determining the credibility of a particular forecast.

[3] The psychology literature uses the terms "preference-consistent" and "preference-inconsistent" to describe information that is congruent or incongruent, respectively, with an individual's preferred belief, regardless of when it is received in a decision-making process (e.g., Ditto and Lopez 1992; Ditto et al. 1998; Ditto et al. 2003). However, for ease of exposition, I will use the terms "favorable" or "unfavorable" to describe the valence of the *initial information* an investor receives subsequent to taking an investment position and "preference-consistent" or "preference-inconsistent" to describe the valence of the *additional information* the investor seeks subsequent to receiving initial information.

The capital markets information environment is unique, in that investors often know—before investing in detailed analysis of a financial report or news article—whether information from a given source supports a particular investment position. For example, websites such as Yahoo! Finance and Dow Jones' Market Watch offer listings of analysts' broad recommendations (e.g., buy, hold, sell) for a given stock. Dow Jones' Market Watch separates analysts' recommendations by upgrades, downgrades, and initiations on its site, providing the name of the brokerage firm, the new recommendation, the old recommendation, and a brief set of comments summarizing the analyst's report. Therefore, investors can quickly determine the flavor of these reports and may, in certain cases, select information based on its level of support for a particular investment position. Moreover, investors' ability to "pre-screen" information is not only limited to analyst reports, but is also inherent in searches for news and popular press articles through databases such as Factiva, ProQuest, and Lexis-Nexis. Federated search engines for business information, such as www.Biznar.com, also allow individuals to search a topic and receive an article title and source.

Given the psychological need to find support for a belief in a prior action when the action is called into question (e.g., Festinger 1957, 1964; Frey 1986) and the nature of today's information environment, I predict that investors seeking support for a previous investment decision after the receipt of unfavorable information will acquire additional information that confirms their position. Specifically, I predict that these investors will choose high-credibility, preference-consistent information over high-credibility, preference-inconsistent information. Moreover, I predict that these individuals, who are in search of information to bolster their belief in their investment position, will forgo a certain amount of credibility to acquire support for that position. On the other hand, I predict that individuals who receive initial favorable information regarding their investment position will experience less of a need to support their position and will seek a balanced set of information from high-credibility sources.

I use a web-based experiment to test my predictions about investors' information acquisition after they receive initial information regarding a previously chosen investment position. Holding constant the initial information given to participants regarding their chosen investment, I manipulate the favorability of that information by manipulating two variables between participants: (1) the participant's directional investment position (long versus short) and (2) the participant's earnings benchmark (high versus low). The randomly assigned investment position (long versus short) introduces a directional preference regarding the firm's earnings outcome, and the randomly assigned earnings benchmark (high versus low) introduces a preference for a particular level of earnings. Those assigned a long (short) position are asked to choose a long (short) investment in one of two firms.

After making their investment choice and learning their earnings benchmark, participants receive initial information about their investment and are provided an opportunity to view additional information in the form of four analyst reports before making an EPS forecast. Two reports are issued by analysts employed by well-known brokerage firms (high-credibility sources) and two by unknown analysts (low-credibility sources). Within each set, one estimates high earnings and the other estimates low earnings. Participants can view as many or as few of the reports as they wish. The primary dependent measure is the time spent viewing each analyst report.

Results generally support my predictions, as I find that the information sought by participants is a function of the favorability of the initial information they receive subsequent to taking an investment position. Holding constant the amount of preference-consistent information offered by high- and low-credibility sources, participants prefer to view high-credibility information. However, participants receiving initial favorable information are more balanced in their acquisition of information based on its preference-consistency than are participants receiving unfavorable information. The majority of high-credibility information chosen by participants receiving unfavorable information is that which is preference-consistent. Additionally, participants receiving initial un-

favorable information prefer to view, at a similar rate, less credible information that is preference-consistent and more credible information that is preference-inconsistent. Consequently, participants' earnings expectations are biased in the direction of the information they choose to view.

These findings contribute to both the accounting and psychology literatures regarding the influence of individuals' preferences on information acquisition and evaluation in three primary ways. First, my study presents a scenario in which the initial information an investor receives regarding a recent investment sets the course of his/her information-acquisition process. Results suggest that an investor receiving unfavorable information regarding his/her investment position will gather additional information in such a way that the resulting information set overly represents information supporting his/her position. Additionally, some of the over-represented information may have less credibility than information that is under-represented. The outcome of this information-acquisition process is a set of information that can result in biased estimates and forgone profits.

Second, the finding that investors are mindful of a source's viewpoint when gathering information has implications for archival researchers investigating market-pricing efficiency (e.g., Diamond and Verrecchia 1987; Figlewski and Webb 1993; Clement and Tse 2003; Gleason and Lee 2003). The results of my study suggest that, under certain circumstances, investors are likely to consider an information source's viewpoint in addition to its credibility in reacting to the information (e.g., Clement and Tse 2003; Gleason and Lee 2003; Mercer 2004; Hutton and Stocken 2009). Additionally, the results present a potential reason for findings that the market's pricing of firms with short-selling constraints, which by definition lack bearish traders, is less efficient than that for firms with short-sale opportunities (Diamond and Verrecchia 1987; Figlewski and Webb 1993).

Third, the study suggests that prior findings in psychology, which find that individuals receiving preference-inconsistent information are sensitive to the credibility of that information and any additional information they might receive, represent a boundary condition (e.g., Lowin 1969; Frey 1981; Ditto et al. 1998). The information offered to participants in these psychology studies is varied between two extreme levels of credibility, resulting in participants choosing high-credibility, preference-inconsistent information over low-credibility, preference-consistent information (Frey 1981). On the other hand, the capital markets information environment offers investors information from various sources representing a wide range of credibility. My study indicates that investors who are seeking support for a desired belief and who are presented with a choice of information from various sources are likely to accept some reduction in information credibility in order to gather support for their position.

The remainder of the paper is organized as follows. Section II develops the theory and hypotheses. The experimental design and experimental results are presented in Sections III and IV, respectively. Section V summarizes and concludes.

II. THEORY AND HYPOTHESES

Source credibility is one way in which an investor's level of uncertainty in information can be reduced (Erdem and Swait 1998). This reasoning is supported by empirical evidence that investors react to information based on attributes of its source (e.g., Jennings 1987; Teoh and Wong 1993; Clement and Tse 2003; Gleason and Lee 2003; Mercer 2004; Rogers and Stocken 2005; Hutton and Stocken 2009), which suggests that investors recognize the importance of forming an accurate estimation of a security's market price. These studies, however, do not consider the current position of an investor. Psychology theory suggests that current investors are likely to find it difficult to be impartial and objective in their search for information regarding a currently held position after receiving information that casts doubt on the profitability of the position (e.g., Festinger

1957, 1964; Frey 1986). The current study examines whether the favorability of initial information an investor receives subsequent to taking an investment position moderates his/her acquisition of additional information regarding that position.

Extant research in accounting and psychology does not provide a clear prediction of investors' information-acquisition process in the course of making judgments and decisions regarding a currently held position. Hales (2007) finds that experimental participants, in the role of investors, interpret analyst forecast information based on their investment position. Specifically, Hales (2007) presents results suggesting that an investor's earnings expectation for an investee firm is more likely to agree with analyst forecast information if that information indicates a high likelihood of a potential gain as compared to that of a loss. However, participants in Hales' (2007) study were not provided with identifying information regarding the reporting analysts' credibility. Research in psychology suggests that an investor receiving unfavorable information regarding his/her investment position will be particularly sensitive to the credibility of the information (Ditto et al. 1998). In turn, s/he would adjust his/her beliefs if the unfavorable information was of high-credibility compared to low-credibility, favorable information.

On the other hand, the psychology research investigating individuals' sensitivity to information based on its credibility and preference-consistency manipulates credibility in an extreme, dichotomous manner (e.g., Lowin 1969; Frey 1981; Ditto et al. 1998). It is unclear how individuals will choose to receive information provided by different sources when the credibility of those sources is varied at more intermediate levels (Lowin 1967). My study seeks a nuanced description of individuals' information-acquisition process when there are preferences to support a desired belief and a more moderate amount of disparity in the various information sources' credibility.

Credibility and Preference Consistency

Cognitive dissonance forms the basis for the effect of directional preferences on individuals' information acquisition (Festinger 1957, 1964). Festinger (1957) describes dissonance as arising when two incompatible sets of knowledge, or cognitions, are held simultaneously. The need to reduce dissonance can lead individuals to be selective in their acquisition of additional information. Because of the opportunity for dissonance to arise upon the initial receipt of information subsequent to taking an investment position, I argue that the favorability of this information regarding the investor's position sets the course of his/her information-acquisition process.

The receipt of information supporting the likely profitability of a recently chosen investment position will corroborate the investor's decision. As such, s/he should experience less need for supportive information, instead seeking a balanced set of information from high-credibility sources regardless of the source's support of the investment position. On the other hand, the receipt of unfavorable information subsequent to an investment choice should arouse dissonance and the investor will, consequently, seek a reduction of the dissonance. One way that individuals can reduce dissonance is by searching for preference-consistent information (Brehm and Cohen 1962).

Research in psychology, however, suggests that investors selectively seeking preference-consistent information will prefer to receive preference-inconsistent information if it is of higher quality than available preference-consistent information (Frey 1981). This finding, along with research suggesting that individuals receiving high-credibility, preference-inconsistent information will update their preferred belief if additional preference-consistent information is low in credibility (Ditto et al. 1998), suggests that individuals seeking additional preference-consistent information are sensitive to source credibility. Therefore, holding constant the amount of preference-consistent information available from high-credibility and low-credibility sources, I expect investors to acquire more high-credibility information than low-credibility information. This prediction is summarized below:

H1: Investors receiving initial favorable or unfavorable information regarding their position will acquire more information from high-credibility sources than from low-credibility sources.

Findings in psychology suggest, however, that given a choice of information from two similarly credible sources, investors will prefer preference-consistent information to preference-inconsistent information (Lowin 1967, 1969; Frey 1986). Such a finding would indicate that an investor is collecting an unbalanced set of information, which could result in the potential overweighting of information supportive of his/her current position in future decision making. Additionally, this result of selective acquisition would have implications for the generality of previous archival findings of investors' reaction to information based on source credibility (e.g., Clement and Tse 2003; Gleason and Lee 2003), as these studies do not control for the current position of investors.

Regardless of an investor's economic incentives, I predict that the need to support a previous investment decision will influence his/her information-acquisition process. I expect that an investor receiving initial favorable information regarding his/her investment position will seek a balanced set of both preference-consistent and preference-inconsistent information from high-credibility sources. On the other hand, I expect an investor for whom initial information is unfavorable regarding his/her position to seek high-credibility information that is preference-consistent over that which is preference-inconsistent. These predictions are summarized below:

H2a: Investors receiving initial favorable information regarding their position will acquire an equal set of preference-consistent and preference-inconsistent information from high-credibility sources.

H2b: Investors receiving initial unfavorable information regarding their position will acquire more preference-consistent information than preference-inconsistent information from high-credibility sources.

Credibility versus Preference Consistency

Previous research examining individuals' selective exposure to information shows that individuals tend to select high-credibility, preference-inconsistent information over low-credibility, preference-consistent information (Frey 1981). This is consistent with findings that information usefulness can influence information search over an individual's need to reduce dissonance (Wicklund and Brehm 1976). The fear of invalidity or a judgmental mistake can lead individuals to seek additional information they view as useful before making a conclusion (Kruglanski and Webster 1996).

The information presented to experimental participants in the psychology studies of selective exposure to information based on source credibility is varied between two extreme levels of credibility (e.g., Lowin 1969; Frey 1981). For example, both Lowin (1969) and Frey (1981) present to experimental participants information that is designed to be either maximally easy or hard to refute. These studies manipulate the messages' ease of refutation by varying the presumed expertise of the information source. Participants were told that the messages were provided by either an expert economist (i.e., low ease of refutation) or by high school sophomores (i.e., high ease of refutation).

The information environment manufactured in the psychology studies described above is very different from the capital markets information environment. With the extensive amount of information available to investors in the marketplace, the sources of information are also vast. Given the availability of information from various sources representing a wide range of credibility, an investor may decide to trade off some amount of credibility in order to gather information sup-

portive of his/her investment position, especially after the receipt of information that casts doubt on the future profitability of his/her position. I predict that investors receiving initial information that is unfavorable regarding their position will not acquire preference-inconsistent information from a high-credibility source at a greater rate than preference-consistent information from a low-credibility source. This prediction is summarized below:

> **H3:** Investors receiving initial unfavorable information regarding their position will *not* acquire more high-credibility, preference-inconsistent information than low-credibility, preference-consistent information.

III. EXPERIMENT

Participants

To test my hypotheses, I conduct a web-based experiment with 92 second-year students from a *BusinessWeek* top-25 M.B.A. program taking on the role of investors. Approximately 58 percent of the participants report that they have made investments in individual firm stocks and 99 percent indicate that they plan to do so in the future. Those who have invested in individual stocks report having purchased a mean (median) of 55 (20) stocks. Additionally, 79 percent of participants report having previously invested in stock mutual funds.

Experimental Procedures

I designed a web-based experimental platform to test whether an investor's need to support an investment decision influences the type of information s/he acquires before completing a basic earnings-forecasting task. Each participant chooses one of two firms in which to invest. A choice is included in the task in order that the decision instills in participants a commitment to their investment position (Brehm and Cohen 1962). In making an investment choice, participants are expected to feel responsible for the future consequences of their decision.

Participants are randomly assigned to either a long position or a short position before choosing a firm in which to invest. Participants view information regarding each of the firms in order to make their choice. The information is not only designed to present the firms as equally attractive, but is also ambiguous regarding the future profitability of the firms.[4] Information presented to participants is held constant across all conditions.[5] After making their choice, participants are randomly assigned either a high or low earnings-per-share benchmark to which actual EPS are compared in determining if they gain or lose from their individual investment.

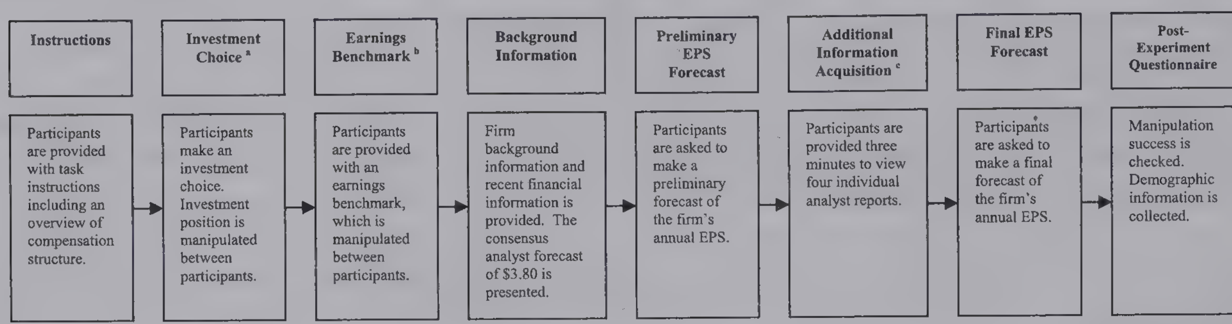
Participants' choice of Firm X or Firm Y has no effect on the information presented subsequently. That is, all participants see the same information after their investment choice, with the exception of the firm's name. Participants receive quarterly earnings information for the past three years, recent historical stock prices over the same period, and the consensus analyst forecast for the upcoming fiscal year, including the range of analysts' individual forecasts. The consensus analyst forecast is either favorable or unfavorable, based on the individual participant's assigned investment position and earnings benchmark. After receiving this information, participants have an opportunity to view any or all of four analysts' reports before making their final EPS estimate. Figure 1 presents a timeline of the experimental task. See the Appendix for an example of the wording of the experimental manipulations.

---

[4] Brehm (1956) finds that post-choice dissonance increases as the attractiveness of the rejected alternative increases. An individual choosing between two attractive alternatives has a choice to make; however, an individual selecting between one attractive alternative and one unattractive alternative will likely feel as if s/he does not have a choice.

[5] This information, including earnings information, is modeled after that of an actual publicly traded firm.

FIGURE 1
Timeline of Experimental Task



[a] Experiment participants assigned to the long (short) investment position condition were asked to choose a long (short) investment in one of two firms. Prior to making their investment choice, all participants viewed the same information regarding the two firms.

[b] Participants were randomly assigned an earnings benchmark of either \$3.76 or \$3.84. Those assigned to the long investment position condition were told that they would receive a higher (lower) payoff when actual earnings were greater (less) than their earnings benchmark. Participants assigned to the short investment position condition were told that they would receive a higher (lower) payoff when actual earnings were less (greater) than their earnings benchmark.

[c] The amount of time a participant spends viewing each of the four analyst reports is the main dependent measure.

Experimental Treatments

I employ a 2 × 2 full-factorial, between-participants design that manipulates (1) the participant's investment position (long versus short) and (2) their earnings benchmark (low versus high).[6] As in Hales (2007), each participant's payoff is either a positive or a negative function of the subject firm's earnings, depending on his/her randomly assigned investment position (long versus short). After participants make their investment choice, they are randomly assigned an earnings-per-share benchmark of either \$3.76 (low) or \$3.84 (high). Those in the long investment position condition experience a higher (lower) payoff when actual earnings are greater (less) than their earnings benchmark. Participants in the short investment position condition experience a higher (lower) payoff when actual earnings are less (greater) than their earnings benchmark. The combination of participants' randomly assigned investment position and earnings benchmark determines whether the initial information received subsequent to their investment choice (i.e., the consensus analyst forecast of \$3.80) is favorable or unfavorable.[7]

After receiving the consensus analyst forecast, participants receive four analysts' forecasts and a link to view each analyst's corresponding report. Two of the individual forecasts, one from a well-known equity research firm and one from an unknown equity research firm, estimate annual

[6] The manipulation of the two independent variables is based on Hales (2007). I am grateful to Jeffrey Hales for his generosity in allowing me access to his experimental instrument.

[7] The consensus analyst forecast of \$3.80 foretells that participants assigned to the long (short) investment position condition and an earnings benchmark of \$3.76 (\$3.84) will experience a gain with regard to their investment position. However, this consensus analyst forecast foretells that participants assigned to the long (short) investment position condition and an earnings benchmark of \$3.84 (\$3.76) will experience a loss based on their investment position.

earnings to be at the top of the range of all analysts' individual forecasts (i.e., \$3.88). The other two forecasts, one from a well-known firm and one from an unknown firm, estimate annual earnings at the bottom of the range of forecasts (i.e., \$3.72). Results from a separate experiment indicate that the well-known firms are expected to provide more credible research than the unknown firms.[8]

The four forecasts are presented in a 2 × 2 matrix rather than a list in order to prevent participants from choosing the analyst reports sequentially. In addition, the computerized program randomizes, across participants, the placement of the forecasts in the matrix. In order to view a particular report, a participant must click on the link associated with that report. Participants are given up to three minutes to view any or all of the four analyst reports; however, participants who spend less than the allotted time are awarded bonus points. Therefore, at any point during these three minutes, participants have the opportunity to stop viewing this information and move forward in the task where they are asked to provide their final EPS forecast.

Dependent Measures

After receiving initial information (i.e., the consensus analyst forecast) that is either favorable or unfavorable regarding their investment position, participants have an opportunity to view four individual analyst reports. I measure the amount of time a participant spends viewing each of the four reports. This allows me to examine whether participants are choosing to view information based on its credibility or its preference consistency. An analyst report will be preference consistent, or will bolster the participant's preferred belief that s/he made a wise investment choice, if it suggests that s/he will experience a gain from the investment (i.e., more optimistic forecasts if a long position is held; less optimistic forecasts if a short position is held). The analyst report will be preference-inconsistent if it casts doubt on the future profitability of the participant's chosen investment position (i.e., less optimistic forecasts if a long position is held; more optimistic forecasts if a short position is held).

Participants' individual viewing times for each report allow me to examine their relative preferences to acquire the information. As the amount of preference-consistent information provided by high- and low-credibility sources is held constant, H1 predicts that all participants will spend more time viewing high-credibility information than low-credibility information. Hypothesis 2a predicts that participants for whom the consensus analyst forecast is favorable will spend equal time viewing the high-credibility, preference-consistent and high-credibility, preference-inconsistent analyst reports. Hypothesis 2b predicts that participants for whom the consensus analyst forecast is unfavorable will spend more time viewing the high-credibility, preference-consistent analyst report than the high-credibility, preference-inconsistent analyst report. Last, H3 predicts that participants for whom the consensus analyst forecast is unfavorable will *not* spend more time viewing the high-credibility, preference-inconsistent analyst report than the low-credibility, preference-consistent analyst report.

[8] In a separate experiment, 41 participants from the same population as that examined in the main experiment rated the credibility of equity research they would expect to receive from each of these four firms. A five-point Likert scale was used with 1 being "not at all credible" and 5 being "highly credible." Mean ratings for the two well-known firms were 4.20 and 4.27. The difference in these ratings is not significant ($t_{40} = 0.83$, $p = 0.41$). The two unknown firms received mean ratings of 3.22 and 3.27. The difference in these ratings is not significant ($t_{40} = 0.36$, $p = 0.72$). However, the average credibility rating for the two well-known firms (mean = 4.23) was significantly greater than that for the unknown firms (mean = 3.24) ($t_{40} = 4.60$, $p < 0.001$). Going forward, I use the labels "low-" and "high-" credibility to connote that the rating of the credibility of the research issued by the two unknown firms was significantly less than the rating of the credibility of the research issued by the two well-known firms. As such, these labels are not meant to convey absolute levels of credibility but, instead, relative levels of credibility.

Compensation

Participants receive compensation based on the number of points earned from four sources. They receive 25 participation points at the beginning of the case. Additionally, they earn points based on the accuracy of their final EPS forecast. If actual earnings equal the participant's forecasted earnings, then s/he earns 25 points; however, one point of the 25 "accuracy points" is deducted for every penny that his/her forecast differs from actual earnings. This component of the compensation scheme provides participants an incentive to make an accurate forecast.

Additionally, participants earn or lose up to 25 points based on the outcome of their investment position. Those holding a long position earn (lose) two points for every penny that actual earnings-per-share exceeds (falls below) their earnings benchmark. Similarly, participants holding a short position earn (lose) two points for every penny that actual earnings-per-share falls below (exceeds) their earnings benchmark. All gains or losses from the participant's investment position are capped at 25 points. This component of the compensation scheme captures investors' incentive to hold a profitable investment.[9]

Lastly, participants have an opportunity to receive bonus points if they do not use the entire three minutes allotted to view the analyst reports. A three-minute timer is provided at the top of the computer screen in order for participants to keep track of their time. The timer clicks down to zero before automatically advancing to the next section of the study where participants are asked to make their final EPS forecast. If participants choose to move forward with 60 seconds or more on the timer, they are awarded ten bonus points. Participants leaving between 30 and 59 seconds on the timer receive five bonus points. These points are added to the participant's total points at the end of the study. This component of the compensation scheme captures the effort expended in information acquisition.[10]

In sum, participants have an opportunity to earn up to 85 points. At the end of the case, each participant's points are converted into U.S. dollars. Participants are guaranteed payment of $5, but could earn up to $25. On average, participants receive compensation equal to $16.[11]

IV. RESULTS

Manipulation Checks

To verify that participants attended to the manipulation of investment position (i.e., long versus short), I asked them in a post-experiment questionnaire to indicate whether they gained or lost points if the firm in which they were invested experienced an increase in earnings.[12] This manipulation was successful: 87 percent (41/47) of participants assigned a long investment position reported that they gained points if earnings rose, while only 7 percent (3/45) of participants

---

[9] To ensure that participants understand the accuracy incentive and profitability incentive components of the compensation scheme, they are required to answer correctly five questions before moving forward in the task.

[10] Studies of selective exposure to information generally force participants to choose between sets of supportive and non-supportive information (e.g., Lowin 1967; Frey and Wicklund 1978; Frey and Stahlberg 1986). Instead of incorporating a similar choice into my study, I implement the time cost. The costliness of analyzing all information should induce choice. Additionally, time spent in an information-processing task is highly correlated with characteristics of effort (Bettman et al. 1990; Sprinkle 2000). Therefore, the time participants spend viewing the analyst reports is a measure of the effort they are willing to expend in gathering additional information.

[11] Compensation was paid at a rate of approximately $0.29 per point earned, with a minimum compensation equal to $5.00. Participants earned an average of 54 points.

[12] Prior to making their investment choice, all participants viewed the same information regarding the two firms. Participants' investment position condition did not influence their investment choice ($\chi^2(1) = 0.69$, $p = 0.41$). In addition, participants' compensation, based in part on their EPS estimate, was not affected by their investment choice ($t_{90} = 0.08$, $p = 0.94$). Participants choosing Firm X earned an average of 53.6 points and received approximately $16 in compensation, while participants choosing Firm Y earned an average of 53.7 points and also received approximately $16 in compensation.

assigned a short investment position used this response. The difference is significant (Fisher's Exact Test; $p < 0.001$). To verify that participants attended to the difference between the consensus analyst forecast (i.e., \$3.80) and their earnings benchmark (i.e., \$3.76 versus \$3.84), I asked them to provide both amounts. Eighty-nine percent (42/47) of those with an earnings benchmark of \$3.76 provided a response that was within \$0.02 of their benchmark. Ninety-six percent (43/45) of those with an earnings benchmark of \$3.84 provided a response that was within \$0.02 of their benchmark. Approximately 97 percent (89/92) of all participants were within \$0.02 of the correct consensus analyst forecast. Most importantly, 94 percent (86/92) of all participants attended to the direction of the consensus analyst forecast in relation to their earnings benchmark. Participants' awareness of the consensus analyst forecast being higher or lower than their earnings benchmark and their responses indicating their understanding that they either gained or lost points when actual earnings increased indicates that the manipulation of the favorability of the consensus analyst forecast was successful.[13]

Mixed-Design ANOVA

As the objective of the study is to examine the effect of the favorability of initial information received subsequent to an investment decision (i.e., the consensus analyst forecast of \$3.80) on investors' viewing time of additional information, I collapse the four experimental conditions into two conditions based on favorability and report the results accordingly. The Favorable condition is comprised of participants assigned to the long (short) investment position with an earnings benchmark of \$3.76 (\$3.84). The Unfavorable condition is comprised of participants assigned to the long (short) investment position with an earnings benchmark of \$3.84 (\$3.76).[14] I then analyze the information using a mixed-design analysis of variance with Favorability as a between-participants variable and the Credibility (high credibility versus low credibility) and Preference Consistency (preference-consistent versus preference-inconsistent) of the analyst reports as two within-participants variables.

Main Effects of Favorability, Credibility, and Preference Consistency on Viewing Time

After receiving initial information regarding their chosen investment positions, participants had the opportunity to view four analyst reports. Based on the participant's assigned experimental condition, each report is one of four types: (1) a high-credibility, preference-consistent report, (2) a high-credibility, preference-inconsistent report, (3) a low-credibility, preference-consistent report, and (4) a low-credibility, preference-inconsistent report. I expect that participants' viewing time of each of these reports will be influenced by the favorability of the initial information they receive regarding their position. Table 1 presents descriptive statistics of participants' viewing time based on Favorability and the type of analyst report (Credibility, Preference Consistency).

The last column of Table 1 shows that participants in the Favorable condition spent, on average, 66.59 seconds viewing the analyst reports, while those assigned to the Unfavorable condition spent an average of 90.48 seconds. The mixed-design ANOVA model presented in Table 2 shows the effect of Favorability on participants' viewing time to be significant ($F_{1,90} = 6.15$, $p = 0.02$). This result is consistent with the findings in psychology that individuals spend more time and effort processing information after the receipt of unfavorable information (e.g., Kruglanski 1980, 1990; Ditto and Lopez 1992).

---

[13] The results reported in the paper are inferentially and statistically similar to the analysis excluding participants who answered the manipulation check questions incorrectly.

[14] I report all significant differences resulting between the two experimental conditions comprising either of the two Favorability conditions.

TABLE 1

Descriptive Statistics for Viewing Time of Analyst Reports[a]

	High-Credibility Reports		Low-Credibility Reports		Collapsed across Credibility		Collapsed Across Preference Consistency		Collapsed across Credibility and Preference Consistency
Favorability of Initial Information[b]	**Preference-Consistent Report[c]**	**Preference-Inconsistent Report[c]**	**Preference-Consistent Report**	**Preference-Inconsistent Report**	**Preference-Consistent Reports**	**Preference-Inconsistent Reports**	**High-Credibility Reports**	**Low-Credibility Reports**	
Favorable	18.93	25.14	12.55	9.98	31.48	35.11	44.07	22.52	66.59
	(14.57)	(22.22)	(14.93)	(11.61)	(25.03)	(25.33)	(31.77)	(20.55)	(44.28)
	n = 44	n = 44	n = 44	n = 44	n = 44	n = 44	n = 44	n = 44	n = 44
Unfavorable	43.02	16.17	18.56	12.73	61.58	28.90	59.19	31.29	90.48
	(45.56)	(15.15)	(17.43)	(16.91)	(50.56)	(22.64)	(44.60)	(26.80)	(47.82)
	n = 48	n = 48	n = 48	n = 48	n = 48	n = 48	n = 48	n = 48	n = 48

The viewing time is in seconds, with standard deviations presented in parentheses.

[a] Participants were given a total of three minutes to view four analyst reports, each containing between 416 and 421 words. Two of the four reports are issued by analysts employed by well-known brokerage firms (high-credibility reports) and two by unknown analysts (low-credibility reports). Within each set, one estimates high earnings and the other estimates low earnings. Participants could view as many or as few of the reports as they wished. The primary dependent measure is the time spent viewing each analyst report.

[b] Participants were randomly assigned to choose either a long or a short investment position and were provided either a $3.76 or a $3.84 earnings benchmark. Subsequent to the participant making an investment choice between two firms and being told their earnings benchmark, all participants were given the same initial information regarding their chosen investment position (i.e., the consensus analyst forecast of $3.80). This information was Favorable for those in the long (short) investment position condition who were assigned a $3.76 ($3.84) earnings benchmark and was Unfavorable for those in the long (short) investment position condition who were assigned a $3.84 ($3.76) earnings benchmark.

[c] The reports of analysts forecasting low (high) earnings should be viewed as preference-consistent by participants assigned to the short (long) investment position condition and as preference-inconsistent by those assigned to the long (short) investment position condition.

TABLE 2

Mixed-Design ANOVA for the Viewing Time of Analyst Reports

	df	SS	MS	F	Two-Tailed p-value
Between-Participants Effects:					
Favorability	1	3275.03	3275.03	6.15	0.02
Error	90	47937.65	532.64		
Within-Participant Effects:					
Credibility	1	14028.97	14028.97	27.17	< 0.001
Credibility × Favorability	1	231.44	231.44	0.45	0.51
Error (Credibility)	90	46464.85	516.28		
Preference Consistency	1	4843.65	4843.65	8.47	< 0.01
Preference Consistency × Favorability	1	7572.31	7572.34	13.25	< 0.001
Error (Preference Consistency)	90	51443.12	571.59		
Credibility × Preference Consistency	1	860.96	860.96	2.16	0.15
Credibility × Preference Consistency × Favorability	1	5094.38	5094.38	12.76	< 0.001
Error (Credibility × Preference Consistency)	90	35935.18	399.28		

The four experimental conditions are collapsed into two conditions based on the Favorability of the initial information (i.e., consensus analyst forecast of $3.80). The Favorable condition is comprised of participants assigned to the long (short) investment position with an earnings benchmark of $3.76 ($3.84). The Unfavorable condition is comprised of participants assigned to the long (short) investment position with an earnings benchmark of $3.84 ($3.76). I analyze the information using a mixed-design ANOVA model with Favorability as a between-participants variable and the Credibility (high credibility versus low credibility) and Preference Consistency (preference-consistent versus preference-inconsistent) of the analyst reports as two within-participants variables.

Holding constant the amount of preference-consistent information offered by high-credibility and low-credibility sources, H1 predicts that all participants will spend a greater amount of time viewing high-credibility reports than low-credibility reports. Consistent with H1, untabulated results show that participants spent an average of 51.96 seconds viewing reports issued by analysts employed by the two high-credibility equity research firms, compared to an average of 27.10 seconds viewing the reports issued by analysts employed by the two low-credibility firms. This difference leads to the significant main effect of Credibility on participants' viewing time of the four reports ($F_{1,90} = 27.17$, $p < 0.001$), as noted in Table 2.

The mixed-design ANOVA in Table 2 also shows a main effect of Preference Consistency on participants' viewing time of the four reports ($F_{1,90} = 8.47$, $p < 0.01$). This main effect, however, is qualified by a two-way interaction of Preference Consistency and Favorability on information viewing times ($F_{1,90} = 13.25$, $p < 0.001$). Compared to participants in the Favorable condition, participants in the Unfavorable condition allocated more of their viewing time to preference-consistent information than to preference-inconsistent information. Participants in the Favorable condition spent approximately the same amount of time viewing the preference-consistent analyst reports (mean = 31.48 seconds) and preference-inconsistent analyst reports (mean = 35.11 seconds) ($t_{43} = 1.01$, $p = 0.32$). However, participants in the Unfavorable condition spent significantly more time viewing the preference-consistent analyst reports (mean = 61.58 seconds) than the preference-inconsistent analyst reports (mean = 28.90 seconds) ($t_{47} = 3.65$, $p = 0.001$).

Credibility and Preference Consistency

Although there is a significant main effect of Credibility on the amount of time participants spent viewing the analyst reports, this main effect is qualified by a significant three-way interac-

tion of Credibility, Preference Consistency, and Favorability ($F_{1,90} = 12.76$, $p < 0.001$). The significant interaction suggests that participants in the Favorable and Unfavorable conditions allocate their viewing time differently across the four analyst reports based on source credibility and the preference-consistency of the report. That is, the two-way interaction of Credibility and Preference Consistency on participants' viewing time varies across the two Favorability conditions. Panels A and B in Figure 2 show graphical presentations of the average viewing times of each of the four analyst reports for participants in the Favorable Condition and Unfavorable Condition, respectively.

Hypothesis 2a predicts that participants in the Favorable condition will be balanced in their acquisition of high-credibility information, choosing to view an equal amount of preference-consistent and preference-inconsistent information. However, H2b predicts that participants in the Unfavorable condition will prefer to view more high-credibility, preference-consistent information than high-credibility, preference-inconsistent information.

Simple effect tests indicate that participants in the Favorable condition spent *more* time viewing the high-credibility, *preference-inconsistent* analyst report (mean = 25.14 seconds) than the high-credibility, preference-consistent analyst report (mean = 18.93 seconds) ($t_{43} = 2.05$, $p = 0.05$). Although this finding does not support H2a, it does not necessarily suggest that participants receiving favorable information regarding their investment position lack objectivity in their acquisition of additional information. Instead, this result is redolent of investors' openness to credible, contrary information (Festinger 1957, 1964; Frey 1986). Additional analysis indicates that participants in the Favorable condition spent equal time viewing the preference-consistent (mean = 12.55 seconds) and preference-inconsistent reports (mean = 9.98 seconds) issued by low-credibility analysts ($t_{43} = 1.00$, $p = 0.33$).

Turning next to participants assigned to the Unfavorable condition, I find that with both high-credibility and low-credibility analyst reports, participants preferred to view the preference-consistent report over the preference-inconsistent report. Specifically, participants in the Unfavorable condition spent, on average, 43.02 seconds viewing the high-credibility, preference-consistent analyst report compared to an average 16.17 seconds viewing the high-credibility, preference-inconsistent analyst report ($t_{47} = 3.63$, $p < 0.001$). This result supports H2b. Similarly, these participants spent, on average, 18.56 seconds viewing the low-credibility, preference-consistent report compared to an average 12.73 seconds viewing the low-credibility, preference-inconsistent report ($t_{47} = 1.88$, $p = 0.07$).

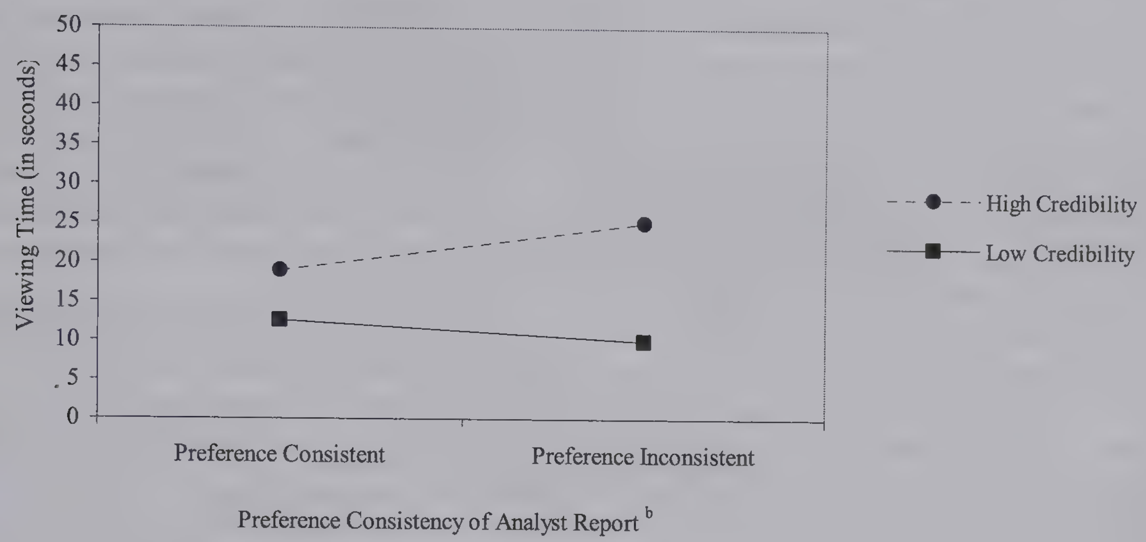
Credibility versus Preference Consistency

Results indicate that participants in the Unfavorable condition seek high-credibility, preference-consistent information over high-credibility, preference-inconsistent information, consistent with H2b. However, I expect these individuals' need to gather information supporting their position to be great enough to counteract the need to gather high-credibility information. Contrary to findings in psychology (e.g., Frey 1981, 1986), H3 predicts that investors receiving initial information that is unfavorable regarding their position will not acquire more high-credibility, preference-inconsistent information than low-credibility, preference-consistent information.

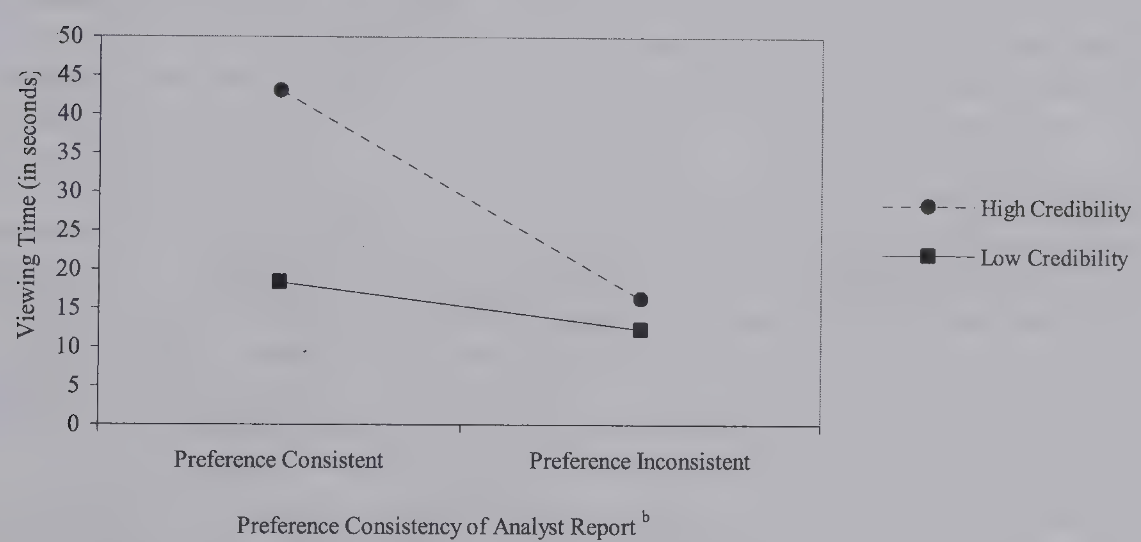
Results support H3. Table 1 shows participants in the Unfavorable condition spent an average of 18.56 seconds viewing the low-credibility, preference-consistent analyst report compared to an average of 16.17 seconds viewing the high-credibility, preference-inconsistent analyst report. The difference in these viewing times is not significantly different ($t_{47} = 0.70$, $p = 0.49$). This result is compelling as previous findings in psychology suggest that individuals receiving unfavorable information prefer additional information that is credible to that which is preference-consistent (e.g., Frey 1981, Ditto et al. 1998). In the current study, however, only participants in the Favorable condition favored credibility over preference-consistency. Putting this result into further

FIGURE 2
Graphical Depiction of Viewing Time of Four Analyst Reports

Panel A: Favorable Condition[a]



Panel B: Unfavorable Condition[a]



[a] The Favorable condition is comprised of participants assigned to the long (short) investment position with an earnings benchmark of \$3.76 (\$3.84). The Unfavorable condition is comprised of participants assigned to the long (short) investment position with an earnings benchmark of \$3.84 (\$3.76).

[b] An analyst report is preference-consistent with a participant's assigned long (short) investment position if the reporting analyst forecasts earnings to be above (below) the participant's earnings benchmark. Conversely, an analyst report is preference-inconsistent with a participant's assigned long (short) investment position if the reporting analyst forecasts earnings to be below (above) the participant's earnings benchmark.

context, participants in the Favorable condition spent an average of 12.55 seconds viewing the low-credibility, preference-consistent analyst report compared to an average of 25.14 seconds viewing the high-credibility, preference-inconsistent analyst report. This difference is highly significant ($t_{43} = 3.82$, $p < 0.001$).[15]

Forecasted Earnings

In order to examine whether selective information acquisition influences subsequent judgments, participants were assigned the task of forecasting annual earnings-per-share for the subject firm. Participants provided a preliminary EPS forecast after receiving the consensus analyst forecast of $3.80 but prior to viewing any of the analyst reports. The average preliminary EPS forecast of participants in each of the four conditions was not different from the consensus analyst forecast of $3.80 (all $p > 0.20$).[16] After having the option to view additional information in the analyst reports, participants provided their final forecasts. Table 3 presents the mean forecasts of participants in each of the four experimental conditions.

The mean forecast of participants assigned to the long investment position was $3.85, while the mean forecast of participants in the short investment position condition was $3.78. The difference of $0.07 is significant ($F_{1,88} = 51.21$, $p < 0.01$). This result replicates the general findings of Hales (2007) that investors' directional preferences regarding an investment position influence their interpretation of information regarding the investment.[17]

Although the average EPS forecast of participants in the Unfavorable, Long condition was greater than the average of those in the Favorable, Long condition, the difference is not significant ($mean_{Unfavorable,\ Long} = \$3.86$; $mean_{Favorable,\ Long} = \$3.84$) ($t_{45} = 1.01$, one-tailed p-value = 0.16). There is a significant difference, however, in the EPS forecasts of participants in the Unfavorable, Short and Favorable, Short conditions ($t_{43} = 1.65$, one-tailed p-value = 0.05), with participants in the Unfavorable, Short condition forecasting EPS to be lower than those in the Favorable, Short condition ($mean_{Unfavorable,\ Short} = \$3.77$; $mean_{Favorable,\ Short} = \$3.79$).[18]

It is interesting to note that while participants in the Unfavorable condition spent more time gathering additional information than did those in the Favorable condition, participants in the

---

15 The results are reported in absolute time (seconds). Due to differences in the total amount of time participants in the Favorable and Unfavorable conditions spent viewing the four reports, I do not test differences in absolute time spent viewing individual reports across the two conditions. However, in calculating the percentage of time participants in both conditions allocated to each report, I find that participants in the Favorable condition spent, on average, 16 percent of their viewing time on the low-credibility, preference-consistent report. This is less than the 21 percent of viewing time, on average, allocated to the low-credibility, preference-consistent report by participants in the Unfavorable condition ($t_{76} = 1.48$; $p < 0.07$, one-tailed). On the other hand, participants in the Favorable condition spent, on average, 39 percent of their viewing time on the high-credibility, preference-inconsistent report, compared to the 21 percent spent, on average, by participants in the Unfavorable condition. This difference is significant ($t_{76} = 3.85$; $p < 0.001$).

16 This suggests that the pattern of participants' additional information search was not due to differing expectations.

17 Hales (2007) predicted and found that individuals assigned to hold a long investment position forecasted earnings to be higher than that forecasted by those assigned to hold a short investment position when evaluating information that implied a loss on their respective investments. However, Hales (2007) found that the earnings forecasts of individuals evaluating information that implied a gain on their respective investments were not significantly different. My findings, however, indicate a difference in the estimates of participants in the Favorable, Long condition and those of participants in the Favorable, Short condition ($t_{42} = 3.96$, $p < 0.01$). A possible reason for this difference is the information participants had an opportunity to view. The experimental design included high-credibility sources forecasting annual earnings at both the top and bottom of the range of individual analysts' forecasts. Therefore, high-credibility information was available for those holding a long (short) position to support a higher (lower) earnings forecast.

18 Forecast errors (i.e., Forecasted EPS – Actual EPS) are presented in Table 3. Actual EPS was $3.81. Participants in the Unfavorable, Short condition had an average forecast error of −$0.04, where participants in the Favorable, Short condition had an average forecast error of −$0.02. The difference in these average forecast errors is significant ($t_{43} = 1.65$, one-tailed p-value = 0.05). Participants in the Unfavorable, Long condition had an average forecast error of $0.05, where those in the Favorable, Long condition had an average forecast error of $0.03. The difference in these average forecast errors is not significant ($t_{45} = 1.01$, one-tailed p-value = 0.16).

TABLE 3
Participants' Final Forecasted Earnings-Per-Share

Earnings Benchmark	Investment Position: Long	Investment Position: Short	Collapsed across Investment Position
\$3.76	\$3.84 (0.04)[a] *\$0.03*[b] n = 23	\$3.77 (0.04) *−\$0.04* n = 24	\$3.81 (0.06) n = 47
\$3.84	\$3.86 (0.05) *\$0.05* n = 24	\$3.79 (0.04) *−\$0.02* n = 21	\$3.83 (0.06) n = 45
Collapsed Across Earnings Benchmark	\$3.85 (0.05) n = 47	\$3.78 (0.04) n = 45	\$3.82 (0.06) n = 92

Shaded cells are Favorable Conditions. Cells not shaded are Unfavorable Conditions.

Participants were given the task of forecasting annual EPS for the subject firm. After receiving background information, including the last three years' earnings information, historical stock prices, and the consensus analyst forecast of annual EPS (\$3.80), participants had an opportunity to view four analyst reports. After viewing this information, participants made their EPS forecast. This task provides an opportunity to examine the differential effects of the information viewed on participants' earnings expectations.

[a] The numbers in parentheses represent standard deviations.

[b] The numbers in italics represent the average forecast error (i.e., Forecast Error = Forecasted EPS – Actual EPS). Actual EPS was \$3.81.

Unfavorable condition provided final EPS forecasts that were, on average, less accurate than those in the Favorable condition. These findings suggest that, in spite of the additional time participants in the Unfavorable condition spent gathering information, their choice to view additional preference-consistent information over a balanced set of both preference-consistent and preference-inconsistent information led them to overestimate (underestimate) the earnings of the firm in which they held the long (short) position.

Lastly, participants' final EPS forecasts are highly correlated with the percentage of total time spent viewing reports issued by analysts forecasting high earnings relative to the consensus ($\rho = 0.56$, $p < 0.01$) and the percentage of total time spent viewing reports issued by analysts forecasting low earnings relative to the consensus ($\rho = -0.57$, $p < 0.01$). Therefore, participants' acquisition of information, which was influenced by the favorability of initial information regarding their individual investment positions, is consequential to their final EPS forecasts. Regardless of the fact that participants' compensation was largely determined by the accuracy of their forecast, participants' final EPS forecasts were biased in a manner consistent with their preferred outcome.

V. CONCLUSION

This study examines whether an investor's psychological preference for information that supports a recent investment decision dominates his/her economic incentive to gather a balanced set of credible information for purposes of future decision making. I argue that the favorability of initial information an investor receives subsequent to taking an investment position sets the course

of his/her information-acquisition process. Given the psychological need to support a belief in a prior action when that action is called into question (Festinger 1957), I predict that investors receiving unfavorable information regarding a recent investment decision will seek additional information that substantiates their decision, even if it requires them to forfeit a certain amount of credibility in that information.

Results from a web-based experiment support my predictions, as I find that the type of information viewed by participants is a function of the favorability of the initial information they received subsequent to making an investment decision. Participants who initially received favorable information chose to view a more balanced set of preference-consistent and preference-inconsistent information compared to participants receiving unfavorable information. Participants receiving initial unfavorable information spent a majority of their viewing time on preference-consistent information, or information that supported the investment position they chose at the beginning of the case. Finally, the participants receiving initial unfavorable information preferred to view at a similar rate less credible information that was preference-consistent and more credible information that was preference-inconsistent.[19] Consequently, participants' earnings expectations were biased in the direction of the information they chose to view.

While archival studies can explore the impact of investors' decisions through stock returns, these studies cannot easily obtain evidence of the processes that precede individuals' investment decisions. An experimental design allows me to manipulate the type of initial information investors receive subsequent to making an investment decision (i.e., favorable versus unfavorable), while controlling the type of additional information available to them, to examine factors that influence information acquisition. In turn, this study allows more nuanced conclusions about investors' use of information relative to extant archival research.

This study contributes to the accounting and psychology literatures examining individuals' search and use of information in three primary ways. First, the results suggest that an investor who initially receives information that casts doubt on a recent investment decision will not gather a balanced set of credible information, but will seek information that bolsters his/her chosen position. The resulting information set overly represents information supporting his/her position. Additionally, some of the over-represented information may have less credibility than information that is under-represented. The outcome of this information-acquisition process is a set of information that can result in biased estimates and forgone profits.

Second, this study contributes to the accounting literature examining the influence of source credibility on investors' use of information (e.g., Teoh and Wong 1993; Clement and Tse 2003; Gleason and Lee 2003; Mercer 2004; Hutton and Stocken 2009). In general, extant research in this area does not consider the current investment position of the individual seeking information. My study suggests that investors are mindful not only of the credentials of the information source, but also of the source's viewpoint. As such, the results have implications for the generality of findings that investors' reaction to information is often based on source credibility (e.g., Teoh and Wong 1993; Clement and Tse 2003; Gleason and Lee 2003). Results also suggest a potential reason for the finding that the market's pricing of firms with short-selling constraints is less efficient than that for firms with short-sale opportunities (Diamond and Verrecchia 1987; Figlewski and Webb 1993). This inefficiency may stem from bullish investors' lack of attention to bad news.

---

[19] Although I find a difference in participants' perceived credibility of the available analyst reports, it is possible that their familiarity with the two well-known brokerage firms led them to rate the credibility of those reports as higher than those issued by two unknown firms. However, given the findings that analysts' forecast accuracy is positively associated with employment at large, high-status brokerage firms (e.g., Clement 1999; Hong and Kubik 2003), I do believe that an analyst's association with a particular employer speaks to his/her credibility.

Last, my study contributes to the psychology literature, as it suggests that prior findings of individuals' sensitivity to information credibility, especially when the information is preference-inconsistent, represent a boundary condition (e.g., Lowin 1969; Frey 1981; Ditto et al. 1998). These psychology studies provide experimental participants with information that varies at extreme levels of credibility. The capital markets information environment, on the other hand, offers investors a vast amount of information from sources representing a wide range of credibility. My findings suggest that investors looking to support a previous investment decision are willing to forgo a certain level of source credibility to find supportive information. However, at the point that the variance in two sources' credibility becomes too extreme, investors will likely choose to acquire information from the more credible source.

APPENDIX

INVESTMENT DECISION

After conducting more than 40 hours of research on the pharmaceutical industry, you concluded that two firms in the industry are **undervalued [overvalued].** Based on your research you would like to take a **long position [short position]** in one of these two stocks.

Note: A **long [short]** investment position is like taking a bet that the stock price will **increase [decrease].**

Upon comparing the two firms, differences between the firms have been narrowed down to a list of factors. These factors, grouped by firm, are as follows:

Firm X

Last three years' growth rates in revenue:

	2006	2005	2004
sales volume	4.80%	5.40%	3.10%
pricing change	1.50%	0.60%	1.00%
currency exchange rate	0.50%	0.70%	3.40%
total growth	6.80%	6.70%	7.50%

Operating Profit Margin last three years:

2006	2005	2004
21%	18%	19%

R&D expense as a % of Total Revenue: 12%

Resignation of CEO, William Lawson, earlier this year

Joint ventures with laboratories in UK, France, Germany, India & China. A licensing agreement is also held with one overseas laboratory.

Increased competition from generic brands in European market

Firm Y

Last three years' growth rates in revenue:

	2006	2005	2004
sales volume	6.30%	7.00%	3.60%
pricing change	1.10%	0.60%	0.60%
currency exchange rate	0.50%	0.50%	4.10%
total growth	7.90%	8.10%	8.30%

Operating Profit Margin last three years:

2006	2005	2004
25%	27%	27%

R&D expense as a % of Total Revenue: 10%

New independent auditor, Ernst & Young LLP; Previous auditor, PricewaterhouseCoopers

Manufacturing facilities located in four Western European countries and one South American country.

2 patent expirations expected in next 18 months; FDA to review application for short extension of patent protection

All other factors you would consider in making this investment were similar across the two firms, including the prices of the respective stocks.

It is important to remember that part of your payment is based on the performance of the investment position you choose. Please choose the firm in which you would like to take a **long [short]** investment position.

INVESTMENT POSITION: HIGHER [LOWER] EARNINGS MEAN MORE POINTS

You chose a long [short] position in Firm X. The price of Firm X stock at the time of your investment was approximately $66 per share. Using the firm's forward price-to-earnings (P/E) ratio of 17.55 [17.19], the implied EPS for the upcoming fiscal year end is $3.76 [$3.84]. Therefore, your investment benchmark is $3.76 [$3.84].

An additional 2 points will be added to your total for every penny by which actual annual EPS is above [below] $3.76 [$3.84], up to 25 points.

Your total points will be reduced by 2 points for every penny by which actual annual EPS is below [above] $3.76 [$3.84], up to 25 points.

If the firm's actual EPS is $3.76 [$3.84], your will neither earn nor lose any points.

In short, higher [lower] earnings leave you with more points. The higher [lower] the firm's relative performance, the better off you are.

REFERENCES

Bettman, J. R., E. J. Johnson, and J. W. Payne. 1990. A componential analysis of cognitive effort in choice. *Organizational Behavior and Human Decision Processes* 45 (1): 111–139.

Brehm, J. W. 1956. Post-decision changes in the desirability of alternatives. *The Journal of Abnormal Psychology and Social Psychology* 52: 384–389.

———, and A. R. Cohen. 1962. *Explorations in Cognitive Dissonance*. New York, NY: Wiley.

Clement, M. B. 1999. Analyst forecast accuracy: Do ability, resources, and portfolio complexity matter? *Journal of Accounting and Economics* 27 (3): 285–303.

———, and S. Y. Tse. 2003. Do investors respond to analysts' forecast revisions as if forecast accuracy is all that matters? *The Accounting Review* 78 (1): 227–249.

Diamond, D. W., and R. E. Verrecchia. 1987. Constraints on short selling and asset price adjustment to private information. *Journal of Financial Economics* 18: 277–311.

Ditto, P. H., and D. F. Lopez. 1992. Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of Personality and Social Psychology* 63: 568–584.

———, G. D. Munro, A. M. Apanovitch, J. A. Scepansky, and L. K. Lockhart. 2003. Spontaneous skepticism: The interplay of motivation and expectation in responses to favorable and unfavorable medical diagnoses. *Personality and Social Psychology Bulletin* 29 (9): 1120–1132.

———, J. A. Scepansky, G. D. Munro, A. M. Apanovitch, and L. K. Lockhart. 1998. Motivated sensitivity to preference-inconsistent information. *Journal of Personality and Social Psychology* 75 (1): 53–69.

Erdem, T., and J. Swait. 1998. Brand equity as a signaling phenomenon. *Journal of Consumer Psychology* 7 (2): 131–157.

Festinger, L. 1957. *A Theory of Cognitive Dissonance*. Stanford, CA: Stanford University Press.

——— 1964. *Conflict, Decision, and Dissonance*. Stanford, CA: Stanford University.

Figlewski, S., and G. P. Webb. 1993. Options, short sales, and market completeness. *The Journal of Finance* 48 (2): 761–777.

Frey, D. 1981. Postdecisional preferences for decision-relevant information as a function of the competence of its source and the degree of familiarity of this information. *Journal of Experimental Social Psychology* 17: 51–67.

——— 1986. Recent research on selective exposure to information. In *Advances in Experimental Social Psychology*, Vol. 19, edited by L. Berkowitz, 41–80. New York, NY: Academic Press.

———, and D. Stahlberg. 1986. Selection of information after receiving more or less reliable self-threatening information. *Personality and Social Psychology Bulletin* 12 (4): 434–441.

———, and R. A. Wicklund. 1978. A clarification of selective exposure. *Journal of Experimental Social Psychology* 14: 132–139.

Gleason, C. A., and C. M. C. Lee. 2003. Analyst forecast revisions and market price discovery. *The Accounting Review* 78 (1): 193–225.

Hales, J. 2007. Directional preferences, information processing, and investors' forecasts of earnings. *Journal of Accounting Research* 45 (3): 607–628.

Hong, H., and J. D. Kubik. 2003. Analyzing the analysts: Career concerns and biased earnings forecasts. *The Journal of Finance* 58 (1): 313–351.

Hutton, A. P., and P. C. Stocken. 2009. Prior forecasting accuracy and investor reaction to management earnings forecasts. Working paper, Boston College.

Jennings, R. 1987. Unsystematic security price movements, management earnings forecasts, and revisions in consensus analyst earnings forecasts. *Journal of Accounting Research* 25 (1): 90–110.

Kruglanski, A. W. 1980. Lay epistemologic—Process and contents: Another look at attribution theory. *Psychological Review* 87 (1): 70–87.

——— 1990. Motivations for judging and knowing: Implications for causal attribution. In *The Handbook of*

Motivation and Cognition: Foundations of Social Behavior, Vol. 2, edited by E. T. Higgins, and R. M. Sorrentino. 333–368. New York, NY: The Guilford Press.

———, and D. M. Webster. 1996. Motivated closing of the mind: "Seizing" and "freezing." *Psychological Review* 103 (2): 263–283.

Leventhal, G. S., and J. W. Brehm. 1962. An experiment on volition of choice. In *Explorations in cognitive dissonance*, edited by Brehm, J. W., and A. R. Cohen. New York, NY: Wiley.

Lowin, A. 1967. Approach and avoidance: Alternate modes of selective exposure to information. *Journal of Personality and Social Psychology* 6 (1): 1–9.

——— 1969. Further evidence for an approach-avoidance interpretation of selective exposure. *Journal of Experimental Social Psychology* 5 (3): 265–271.

Mercer, M. 2004. How do investors assess the credibility of management disclosures? *Accounting Horizons* 18 (3): 185–196.

Rogers, J. L., and P. C. Stocken. 2005. Credibility of management forecasts. *The Accounting Review* 80 (4): 1233–1260.

Sprinkle, G. B. 2000. The effect of incentive contracts on learning and performance. *The Accounting Review* 75 (3): 299–326.

Stickel, S. E. 1992. Reputation and performance among security analysts. *The Journal of Finance* 47 (5): 1811–1836.

Teoh, S. H., and T. J. Wong. 1993. Perceived auditor quality and the earnings response coefficient. *The Accounting Review* 68 (2): 346–366.

Wicklund, R. A., and J. W. Brehm. 1976. *Perspectives on Cognitive Dissonance*. Hillsdale, NJ: Lawrence Erlbaum Associates.

THE ACCOUNTING REVIEW
Vol. 86, No. 1
2011
pp. 23–57

American Accounting Association
DOI: 10.2308/accr.00000008

The Spread of Aggressive Corporate Tax Reporting: A Detailed Examination of the Corporate-Owned Life Insurance Shelter

Jennifer L. Brown
Arizona State University

ABSTRACT: This study investigates the spread of aggressive corporate tax reporting by modeling a firm's decision to adopt the corporate-owned life insurance (COLI) shelter. Prior studies identify firm characteristics associated with aggressive tax reporting (Desai and Dharmapala 2006; Frank et al. 2009) and tax shelter participation (Wilson 2009; Lisowsky 2010). This study examines whether social environment factors explain the pattern of tax shelter adoption. Building on theory related to the diffusion of innovations and institutional isomorphism, I hypothesize direct and indirect ties between prior and potential shelter adopters influence the spread of shelter use. I find that network ties via board interlocks increase the likelihood of adopting the COLI shelter. I also find weak evidence that COLI use spreads geographically. However, I find no evidence that the spread of COLI use is concentrated among a particular set of audit firms or industries.

Keywords: *tax shelters; tax aggressiveness; corporate reporting; diffusion; institutional isomorphism; board interlocks.*

Data Availability: *All data are publicly available from sources identified in the text.*

I. INTRODUCTION

During the 1990s, the corporate tax shelter industry boomed, drawing coverage from the financial press and a full-scale crackdown by the U.S. Treasury. In 1998, *Forbes* magazine described a new breed of tax shelters in its cover story, "The Hustling of X-Rated Shelters" (Novack and Saunders 1998). The following year, the U.S. Department of the Treasury (1999) released a 164-page report urging the adoption of numerous legislative measures to curb the growing tax shelter problem. However, to date we have limited evidence on the factors that affect a firm's decision to adopt a tax shelter (Wilson 2009; Lisowsky 2010).

This paper is based on my dissertation, and I appreciate the helpful suggestions and insights of my dissertation committee: John Robinson (chair), Steve Limberg, Linda Krull, Rob Williams, and Jay Hartzell. I also thank Lil Mills, Stephanie Sikes, Steve Kaplan, Katharine Drake, and workshop participants at Arizona State University, Emory University, Florida State University, George Mason University, Rice University, Texas Christian University, University of Illinois at Urbana–Champaign, The University of Texas at Austin, and University of Washington for valuable comments. I gratefully acknowledge the financial support of the Deloitte & Touche Foundation.

Editor's note: Accepted by Thomas Omer.

Submitted: June 2008
Accepted: June 2010
Published Online: January 2011

Like the Enron-era wave of financial accounting scandals, the most recent corporate tax shelter boom alarmed regulators because it signaled a shift in the corporate norm toward aggressive reporting. In outlining its concern over the proliferation of tax shelters, the U.S. Treasury points to the following quote from the New York Bar Association:

> The constant promotion of these frequently artificial transactions breeds significant disrespect for the tax system, encouraging responsible corporate taxpayers to expect this type of activity to be the norm, and to follow the lead of other taxpayers who have engaged in tax advantaged transactions.
> U.S. Treasury (1999, 3)

This quote motivates the question, "How does a corporate practice such as the use of tax shelters spread?" and suggests that later shelter adopters imitate early shelter adopters because they see tax shelter use as the new norm. I investigate the spread of aggressive corporate tax reporting by examining investments in a specific tax shelter: corporate-owned life insurance (COLI).[1] Drawing on theories from the innovation-diffusion literature, I test whether factors related to a firm's social environment help to explain tax shelter participation.

Rogers (2003, 11) defines diffusion as "the process by which (1) an *innovation* (2) is *communicated* through certain *channels* (3) *over time* (4) among the members of a *social system*" (emphasis in the original). Because firms' strategic choices are similar to innovations, researchers have looked to innovation-diffusion theory for insights on how and why certain firm practices spread (such as poison pills [Davis 1991], investor relations departments [Rao and Sivakumar 1999], multi-divisional form [Palmer et al. 1993], golden parachutes [Davis and Greve 1997], options back-dating [Bizjak et al. 2009], and private equity targets [Stuart and Yim 2010]). The decision to adopt an innovation is heavily influenced by the decision-maker's broader social environment (Rogers 2003). Before adopting an innovation, decision-makers must have knowledge of the new strategy—knowledge often gained through direct contacts between peers or indirect contacts with other agents of change within the social environment. Once aware of the innovation, potential adopters seek to reduce the inherent uncertainty that surrounds the implementation of novel ideas by looking to the experiences of their peers.

The importance of obtaining information on new strategies vicariously by observing prior adopters links diffusion theory with a separate, but related field of study: institutional isomorphism. Institutional isomorphism is concerned with the homogenization of organizational behavior, when, why and how firms mimic one another. When faced with uncertainty, organizations economize on search costs and imitate the behavior of other organizations (Cyert and March 1963). Mimetic, or imitative, organizational change can be thought of as a contagion process that spreads fashionable practices from one firm to another (Haveman 1993). Theory on the diffusion of innovations and institutional isomorphism suggests that firms imitate their peers in the adoption of new strategies (for example, see Haveman 1993; Haunschild 1993; Rao and Sivakumar 1999). I extend the application of innovation-diffusion theory to examine the spread of aggressive tax reporting, focusing on the innovative COLI shelter developed in response to the Tax Reform Act of 1986 (TRA86).

First, I analyze the pattern and timing of COLI shelter adoptions. I find that firms' adoption decisions appear to be influenced by prior COLI adopters and by the possibility of changes to the tax law. Then, using a matched control sample design, I estimate a cross-sectional logit model of tax shelter adoption that includes four social environment factors suggested by prior studies on the diffusion of strategic corporate practices: (1) direct contacts between prior and potential adopters

---

[1] I describe the development of the corporate-owned life insurance shelter in Section II. Unless otherwise indicated by context, I use the term "COLI" to refer specifically to the COLI shelter, not corporate-owned life insurance in general.

(proxied by board ties); (2) indirect ties through shared professionals (proxied by auditor ties); and imitation of (3) structurally equivalent (industry) and (4) geographically proximate peers.

I find that direct ties with prior COLI adopters via board interlocks increase the probability that a firm adopts COLI. I also find some evidence that COLI adoptions spread geographically. However, although auditors were targeted in the public press for their role in the tax shelter boom, I find no evidence that COLI adoptions spread via shared auditors. Similarly, I find no evidence that COLI adoptions spread along industry lines. Finally, I test the robustness of these findings using event history analysis and a hazard model that captures both the likelihood of adoption and the rate at which a firm adopts, and find similar results with respect to board interlocks.

This study contributes to recent literature on aggressive corporate tax reporting by examining the spread of a specific tax shelter, COLI. Prior studies have uncovered various firm characteristics associated with aggressive corporate tax reporting (Desai and Dharmapala 2006; Frank et al. 2009) and tax shelter use (Wilson 2009; Lisowsky 2010). Although the results of these studies help to explain cross-sectional variation in aggressive reporting behavior and help researchers identify which firms were likely engaged in aggressive behavior, they do not offer much insight into the wave of aggressive reporting that prompted the U.S. Treasury's report in 1999. Firm behavior is clearly related to firm characteristics, but it is also embedded in larger social structures and culture. This study examines the impact of those social structures and thus contributes to an understanding of the macro-level behavior that took place during the corporate tax shelter boom.

Section II details the development and mechanics of the COLI tax shelter and the identification of COLI adopters examined in this study. Section III discusses COLI in the context of innovation-diffusion theory. Section IV develops a cross-sectional model of COLI adoption and discusses the estimation results. Section V describes the event history analysis and presents the hazard model estimation results. Section VI concludes.

II. BROAD-BASED LEVERAGED COLI

Development of the COLI Shelter

The COLI shelter can be described best as a tax arbitrage transaction: the taxpayer finances the purchase of an asset that produces tax-exempt income, a cash value life insurance policy, using proceeds from a loan that produces tax-deductible interest expense.[2] The cash value policy includes a death benefit and a savings component, both of which are accorded preferential tax treatment. Unlike other interest income, which is taxed as earned, the interest credited to the cash value of a policy (the inside buildup) is excluded from gross income (IRC §72).[3] Moreover, a taxpayer can access the cash value of his/her policy while still preserving the deferral benefit by borrowing against the policy because loans secured by life insurance contracts are not treated as taxable distributions (IRC §72).

Over time, Congress has tried to limit life-insurance-related tax arbitrage by narrowing the statutory definition of life insurance. Concerned about the excessive amount of "key-man" life insurance being purchased by firms, Congress limited interest deductions on COLI borrowings by capping qualified policy loans at $50,000 per insured life (TRA86). Instead of abandoning the market for tax arbitrage after TRA86, insurance industry entrepreneurs responded by offering a

---

[2] Since payouts upon the death of the insured are generally tax-exempt to the beneficiary, the premium expense related to the production of that income is not deductible, hence the need for the arbitrage transaction.

[3] Distributions of the cash surrender value are generally treated first as a tax-free recovery of basis, and only result in includible income when the amounts distributed exceed the taxpayer's investment or basis in the policy (IRC §72(e)). Thus, even taxpayers who withdraw the cash surrender value of a policy can still enjoy a significant deferral benefit.

new product: broad-based, leveraged COLI (BBCOLI). BBCOLI uses volume, covering thousands of a company's rank-and-file employees, to make up for the arbitrage opportunity denied by the $50,000 per insured policy loan cap.[4]

BBCOLI provides a good setting for studying the spread of aggressive corporate tax reporting. Although defining exactly what constitutes aggressive corporate tax reporting is challenging, BBCOLI is generally classified by accounting researchers and legal scholars as a corporate tax shelter.[5] Furthermore, details from the Camelot case, *In re CM Holdings, Inc.* (254 B.R. 578, D. Del. 2000), reveal exactly when this innovative tax shelter was developed, a critical element in studying the diffusion of a practice. Last, BBCOLI is a good setting because changes to the law in 1996, government court victories, and the implementation of an IRS amnesty/settlement program effectively shut down the COLI shelter, forcing COLI investors to unwind their investments and bringing COLI investors out of hiding.[6] Thus, COLI represents a chance to identify and study firms that invest in tax shelters, both those that were caught and those that voluntarily came forward.

Identifying COLI Shelter Participants

I use Lexis-Nexis to search 10-Ks, news articles, and press releases for evidence of COLI activity using the following string: *(company or corporate) w/5 owned w/5 life w/5 insurance w/10 (tax or taxes)*. This search captures two main groups of firms: (1) those for which involvement in the COLI shelter is revealed *ex post* because of an IRS dispute, IRS amnesty settlement, or other legal dispute, and (2) those that openly disclose their COLI activity in their financial statements during active COLI shelter years.[7] Owens & Minor is an example of the first group; footnote disclosures in 2000 and 2001 discuss the IRS Notice the firm received regarding COLI and the reserve charge booked in light of the IRS's COLI court victories. Marriott is an example of the second group, as shown in the following excerpts from the firm's 1995 annual report. The top excerpt comes from Management's Discussion and Analysis (MD&A), and the bottom excerpt is the reconciliation of the firm's effective tax rate (ETR) to the statutory tax rate provided in the notes to the financial statements.

> The Company's effective income tax rate declined to 40.0 percent from 41.5 percent in the preceding year, despite expiration of federal jobs tax credit programs, due to the impact of the Company's **corporate-owned life insurance** program and certain other investments.

---

[4] Winn-Dixie covered nearly all 36,000 of its employees while Wal-Mart allegedly covered 350,000 of its employees (*Winn-Dixie Stores*, 113 T. C. 54 (1999); *Rice v. Wal-Mart Stores, Inc.* 12 F. Supp. 2d 1207).

[5] COLI is cited as a classic example of the recent breed of corporate tax shelters (Bankman 2004; Eustice 2002; Gergen 2002) and is included in analyses by Graham and Tucker (2006), Wilson (2009), and Lisowsky (2010).

[6] In 1996, Congress further restricted COLI arbitrage, limiting a corporation's interest deductions to those on policy loans for a maximum of 20 individuals (Health Insurance Portability and Accountability Act, HIPAA 1996). The IRS successfully litigated three COLI cases, *Winn-Dixie* (113 T. C. 254), *C.M. Holdings* (254 B.R. 578), and *AEP* (136 F. Supp 2d 762), and implemented an amnesty/settlement program for COLI transactions.

[7] Tax shelters are usually characterized by secrecy and using a sample of firms that openly disclose their activities in their financial statements is likely to call into question whether these firms truly represent typical tax shelter users. However, Bankman (2002) suggests that firms that think their shelter activities would need a cloak are less likely to engage in tax shelters, while those that decide to adopt tax shelters have convinced themselves that their shelter activity meets a littoral reading of the tax law.

	1995	1994	1993
U.S. statutory tax rate	35.0%	35.0%	35.0%
State income taxes, net of U.S. tax benefit	5.0	5.0	6.0
Corporate-owned life insurance	**(1.2)**	—	—
Tax credits	(1.4)	(0.7)	(0.8)
Other, Net	2.6	2.2	2.0
	40.0%	41.5%	42.2%

My initial text-string search results in 213 possible COLI adopters. I exclude 28 life insurance companies because these firms serve as the other party in the COLI tax shelter transaction. Bank-owned life insurance, BOLI, is similar to COLI, but since the legislative history surrounding limitations on BOLI is different, I exclude 32 banks from the sample of COLI shelter participants.[8] After excluding banks and insurance firms, 153 potential COLI shelter participants remain.

Firms often purchase COLI for legitimate business purposes, including insuring against the loss of key executives and providing a perk for outside directors. To ensure that I only include firms that participate in the tax shelter variety of COLI, I examine the remaining 153 firms in detail to confirm participation in BBCOLI. I assume that firms whose COLI programs give rise to tax benefits material enough to warrant a line-item on the tax rate reconciliation are involved in BBCOLI and include these firms in the shelter sample. I identify additional COLI shelter participants based on news articles, court cases, and financial statement details regarding COLI-related IRS disputes or settlements. After examining the 153 firms from my initial search in detail, I identify 43 firms as COLI shelter participants.[9]

Estimate of COLI Tax Savings

Table 1 provides various estimates of the magnitude of COLI tax savings for the sample firms. Twenty-seven (62.8 percent) of the sample firms provide detailed information on their COLI tax savings through their tax footnote rate reconciliations. Using these details, I report estimates of the average annual and total tax savings under the columns labeled "Rate Reconciliation."[10] On average, over the duration of COLI use, firms report saving $26 million dollars in taxes, equivalent to 2.9 percent of pre-tax book income.

For firms that disclose the dollar amount of their COLI-related IRS settlement, I report the amount and year of settlement under the columns headed, "Settlement." The mean (median) settlement amount is $50.8 million ($38 million), nearly twice the estimated amount of tax savings reported in firms' rate reconciliations. The last two columns report estimated COLI savings derived from a variety of other sources, including court cases and footnote disclosures of COLI-related reserve charges. The mean COLI tax savings estimated from these sources is $46.8 million. In 2000, the staff of the Joint Committee on Taxation estimated that there were 100 COLI cases involving nearly $6 billion dollars in taxes (Paull 2000). Multiplying the highest average tax savings estimate, $50.8 million, times the 43 firms in the sample yields an aggregate tax savings estimate of more than $2 billion, suggesting my sample accounts for roughly one-third of the COLI tax shelter market.

---

8 Neither the economic sham arguments used in court to defeat leveraged COLI nor the 1996 HIPAA restrictions on leveraged COLI directly apply to BOLI because banks generally do not borrow against the policies to fund the premium payments.

9 My sample includes all firms whose shelter activity is observable. Tax shelter activity is observable if either (1) the firm is a disclosing firm, (2) the firm is a non-disclosing firm whose activity is revealed *ex post* when detected by regulators, or (3) the firm is a non-disclosing firm whose activity is revealed *ex post* when the firm comes forward for the COLI amnesty program. Tax shelter activity is generally not observable for non-disclosing firms whose activities are not detected by regulators.

10 Because the dollar-effect of rate reconciliation line items is harder to interpret in years when pre-tax book income is negative, I only include COLI tax savings amounts for years with positive pre-tax book income.

TABLE 1

Estimates of COLI Tax Savings

Obs.	Company	Rate Reconciliation					Settlement		Other	
		Year of Adoption	No. of Years	Annual Savings	Total Savings	Total as a Percent of PTBI	Amount	Year Disclosed	Amount	Source
1	Advanced Telecommunications	1991	1	$585	$585	1.96%				
2	American Electric Power	1990	7	19,418	135,926	2.49%			$25,000	136 F. Supp 2d 762[a]
3	American Greetings	1990	10	9,208	92,085	4.50%			143,000	2001 Q4 reserve charge
4	Ball Corp.	1987	11	3,616	39,775	4.38%				
5	Bassett Furniture Industries	1994	5	726	3,631	2.70%			8,000	1998 fn.: total tax effect
6	Brush Engineered Materials	1990	8	715	5,722	3.33%				
7	Carpenter Technology Corp.	1994	NA							
8	Charming Shoppes Inc.	1993					$18,477	$2,004		
9	Commercial Intertech	1992	7	1,118	7,824	3.48%				
10	Diebold Inc.	1987	12	2,428	29,131	2.78%	16,040	2002		
11	Donnelly Corp.	1990	9	514	4,628	4.72%	217,000	2002*	186,000	1999 fn.: total tax effect
12	Dow Chemical	1989							11,247	250 F. Supp 2d 748
13	Eaton Corp.	1993	NA							
14	Fina Inc.	1989	9	3,043	27,385	2.23%				
15	GATX Corp.	1988	10	4,007	40,071	3.27%				
16	Goodrich Corp.	1990	7	1,565	10,956	0.91%				
17	Grace (W.R.) & Co.	1989					57,500	2005		
18	Great Plains Energy Inc.	1994					12,700	2002*		
19	Hershey Co.	1989	5	3,043	15,216	0.79%			61,200	2000 fn.: payment to IRS
20	Hillenbrand Industries	1990	NA							
21	Hovnanian Entrprs Inc.	1993	6	675	4,050	2.20%				
22	Hubbell Inc.	1993	NA							
23	IKON Office/ALCO Standard	1994							10,000	2001 reserve charge
24	Manor Care Inc.	1993	5	4,254	21,268	5.69%	38,000	2001*		
25	Marriott International Inc.	1995	4	4,438	17,753	0.85%				

(continued on next page)

TABLE 1 (continued)

		Rate Reconciliation					Settlement		Other	
Obs.	Company	Year of Adoption	No. of Years	Annual Savings	Total Savings	Total as a Percent of PTBI	Amount	Year Disclosed	Amount	Source
26	Media General	1990	9	1,525	13,727	1.82%				
27	Miller (Herman) Inc.	1994	3	1,867	5,602	2.80%	22,200	2003*		
28	Norfolk Southern Corp.	1989	10	10,300	103,000	1.15%				
29	OLIN Corp.	1992					41,000	2004*		
30	Owens & Minor Inc.	1995							7,200	2001 reserve charge
31	Pitney Bowes Inc.	1994	5	7,249	36,245	1.02%			76,000	2005/2006 reserve charge
32	PPG Industries Inc.	1994	NA							
33	Rohr Inc.	1994	3	223	669	3.34%				
34	Ruddick Corp.	1993	7	2,648	18,535	3.85%	20,000	2001*		
35	Sherwin-Williams Co.	1994	NA							
36	Sonoco Products Co.	1993	5	5,604	28,019	2.05%			23,300	2000/2001 reserve charge
37	Stride Rite Corp.	1993	5	1,046	5,228	2.63%				
38	Swank Inc.	1986	5	326	1,632	9.64%				
39	Trans World Entmt. Corp.	1991							5,534	254 B.R. 578
40	Unifirst Corp.	1994	5	716	3,579	1.82%			5,500	1999 reserve charge
41	Wal-Mart Stores Inc.	1993	NA							
42	Winn-Dixie Stores Inc.	1993	5	6,124	30,619	1.77%	52,000	2003*	1,599	113 T. C. 254[b]
43	Xcel Energy Inc.	1993					64,400	2007*		
	MEAN			3,592	26,032	2.90%	50,847		46,832	
	MEDIAN			2,428	15,216	2.63%	38,000		17,274	

(continued on next page)

TABLE 1 (continued)

* Indicates that the dollar amount of year disclosed includes interest on the tax deficiency.
[a] Amount is the tax deficiency for the 1996 tax year only.
[b] Amount is the tax deficiency for the 1993 tax year only.

This table presents estimates of COLI-generated tax savings based on details from court cases and firms' financial statements. No. of Years equals the number of years prior to 1998 with positive pre-tax book income (PTBI) for which the firm reports COLI savings on the rate reconciliation. Total Savings equals actual annual savings reported on the rate reconciliation summed across those years. Annual Savings equals total savings divided by No. of Years. Total as a Percent of PTBI equals total savings divided by the sum of PTBI across related years. Settlement amounts are identified from financial statement disclosures. The final two columns present estimates of COLI tax savings from a variety of other sources. All dollar amounts are in thousands.

III. THE DIFFUSION OF COLI

Pattern of COLI Adoption

I examine the pattern of COLI adoptions over time. There are three basic models of diffusion: the internal influence model, the external influence model, and the mixed influence model, which allows for both internal and external influences (Mahajan and Peterson 1985).[11] Imitation, suggested by the U.S. Treasury (1999) as a factor in the spread of shelter use, is characteristic of the internal influence model, which predicts that the rate at which an innovation is adopted is a function of the cumulative number of prior adopters:

$$\frac{dN(t)}{dt} = x(t) = fn(N(t-1)), \tag{1}$$

where $N(t)$ is the cumulative number of adopters at time t, and $x(t)$ is the number of new adopters at time t. Under the internal-influence model, the rate of adoption, $\frac{dN(t)}{dt}$, plotted against time produces a bell-shaped curve and the cumulative diffusion curve is S-shaped (Mahajan and Peterson 1985; Rogers 2003).

In contrast, external models assume no relationship between $x(t)$ and $N(t-1)$—adoption is driven solely by an external source like mass media or a change agent. Under the external-influence model, the cumulative number of adopters increases over time, but at a (constant) decreasing rate (Mahajan and Peterson 1985). COLI was developed by an insurance entrepreneur and marketed by insurance brokers, and these tax shelter promoters, acting as external agents of change, obviously had an impact on the spread of COLI use. Although court documents reveal details of the relationship between shelter promoters and a few COLI adopters, the extent of contact between tax shelter promoters and all potential COLI adopters is unobservable. Therefore, I generally assume that the influence of external change agents, in the form of tax shelter promoters, is constant across the sample of potential COLI adopters examined in this study.[12]

I examine the pattern of COLI adoption by using data from firms' rate reconciliations and disclosures regarding IRS disputes and settlements to estimate t, the year of COLI adoption.[13] As Figure 1 shows, the cumulative number of COLI adopters generally increases from 1986 to 1990, levels off in 1991 and 1992, and increases again in 1993. The increases in new adoptions in 1990 and 1993 can be explained by the legislative history of anti-COLI measures. By 1990, Congress and the U.S. Treasury were aware of COLI abuses (U.S. Department of the Treasury 1990), and in 1991 anti-COLI legislation was proposed in the Senate. Fearing new restrictions on COLI use, but expecting that any existing COLI programs would be grandfathered, firms that were considering adoption in early 1990 rushed to complete their COLI deals (Gergen 2002). After Congress failed to enact the proposed restrictions in 1991 and 1992, the rate of COLI adoptions surged in 1993.

While Figure 1 indicates that the rate of COLI adoptions over time is likely related to the history of proposed legislative measures to curb COLI use, the observed rate of COLI adoptions

---

[11] Importantly, these models assume that an innovation is legal and do not consider the role of an external regulatory body with the power to curb or halt the rate of adoption. Gergen (2002) analytically models how the risk of detection can incentivize tax shelter promoters and tax shelter users to limit the volume of a shelter's distribution. However, this study is the first to empirically model the spread of shelter use.

[12] Shelter promoters could have systematically targeted firms with the same firm characteristics I model as predictors of adoption, giving rise to an endogeneity problem. However, a systematic bias in the type of firm targeted by shelter promoters should bias against findings related to the social environment factors included in the model.

[13] Since managers can exercise considerable discretion in the amount of detail disclosed in the firm's rate reconciliation, using rate reconciliations may lead to mismeasurement of t, but any mismeasurement should bias against finding results.

FIGURE 1
Observed Pattern of Cumulative COLI Adoptions

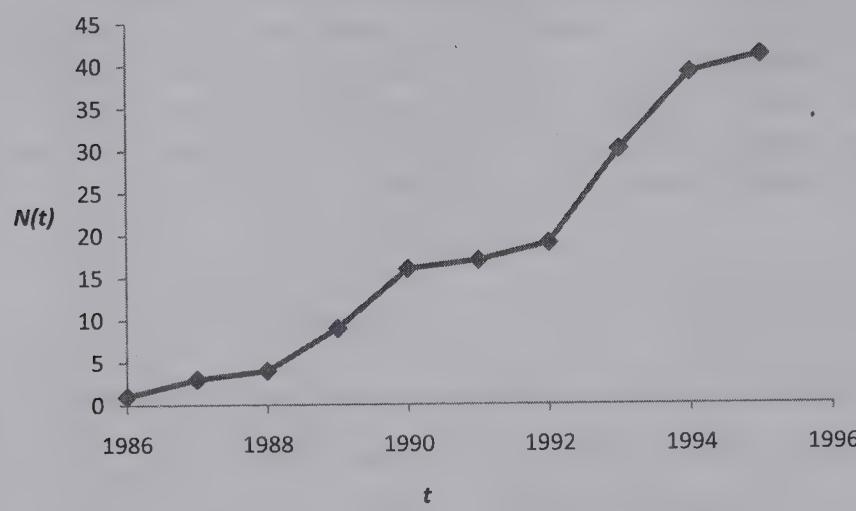


Figure 1 presents the pattern of cumulative COLI adoptions, $N(t)$, observed in the COLI sample.

does not map clearly into the pattern predicted by either a pure internal or pure external source diffusion model. Below, I present four possible mechanisms of diffusion and test whether each helps to explain the spread of COLI use.

Sources of Internal Influence

Innovation-diffusion is linked to mimetic isomorphism, the tendency of firms to imitate one another, because, in an attempt to gain information and reduce the inherent uncertainty surrounding an innovation, potential adopters rely heavily on the experiences of prior adopters. The importance of evaluating an innovation vicariously though the trials of prior adopters is consistent with anecdotes from shelter participants reported by Bankman (1999, 1781): "Most companies are reluctant to be the first purchasers, preferring to purchase a shelter that has been vetted by others." Under frequency-based imitation, as more and more firms adopt a practice, the practice gains legitimacy and eventually becomes taken for granted (March 1981).

Outcome-imitation theorizes that firms adopt strategies when they observe improvements in the performance of peer firms using those strategies (Haunschild and Miner 1997). Abrahamson and Rosenkopf (1993) posit that pressures on firms arising from the threat of lost competitive advantage produce competitive bandwagons, even when the returns to an innovation are unclear. As the number of adopters in a group increases, non-adopters face the risk of falling farther below average if the innovation succeeds. Outcome-imitation is consistent with the U.S. Treasury's (1999) report, which identifies pressure to keep ETRs low and in line with competitors as a driving force behind the tax shelter boom. Both frequency-based imitation and outcome-based imitation imply that the prevalence of a practice within a community will impact the rate of adoption.

Cohesion

Cohesion models of imitation focus on personal contacts between prior adopters and potential adopters. In order to adopt an innovative practice, potential adopters must be aware of the innovation's existence *and* be persuaded to adopt it. A potential adopter can learn about a tax shelter

through its network contacts. More importantly, network contacts with prior adopters can help potential adopters evaluate the benefits and risks associated with a shelter, and therefore impact the persuasion stage of the innovation-diffusion process. Unfortunately, network contacts are often difficult to identify and measure. Tax shelter participants can be connected through common professional association memberships, common board members, common legal advisors, common tax consultants, common auditors, or others.

Board interlocks have been shown to affect the diffusion of a variety of corporate practices (e.g., Davis 1991; Mizruchi 1992; Haunschild 1993; Palmer et al. 1993; Bizjak et al. 2009). The estimates provided in Table 1 suggest that COLI tax savings were significant, and accounts of the negotiations between insurance brokers and firms surrounding the initiation of COLI programs document top management's involvement in the decision to adopt COLI.[14] As such, it is reasonable to assume that board members were aware of COLI and could have influenced the rate of COLI adoptions through their ties to other boards.[15] Even absent this direct type of influence, board interlocks proxy as a general indicator that two firms are socially connected (Strang and Soule 1998). Therefore, I hypothesize:

H1: Ties to prior COLI adopters via board interlocks will increase the likelihood that a firm adopts the COLI shelter.

Professionals, like lawyers, consultants, and accountants, can act as external change agents, influencing the adoption of a new corporate practice by bringing the innovation to their clients. For example, court documents indicate that AEP initially considered COLI at the suggestion of their accountants, Deloitte, Haskins, and Sells (*AEP, Inc. v. U.S.*, 136 F. Supp. 2d 762). Professionals can also play a role in firms' mimetic behavior, acting as the indirect tie connecting prior and potential adopters (DiMaggio and Powell 1983). Haunschild (1994) finds that professional firms act as a conduit for the spread of acquisition premium information. The uncertainty engendered with considering an activity of questionable legitimacy, like tax sheltering, can make firms more likely to rely on information gathered through professional firms, if only to legitimate and rationalize the decision to engage in the activity (Pfeffer 1981).[16]

Regulators have focused on the role of public accountants in the spread of the tax shelter industry, criticizing that when a firm advises its audit client on a tax shelter transaction, the firm essentially audits its own work (U.S. Senate Permanent Subcommittee on Investigations 2005). Because regulators investigating the tax shelter industry have placed a spotlight on auditors and because auditors are a natural information conduit between prior and potential shelter adopters, I hypothesize:

H2: Ties to prior COLI adopters via shared auditors will increase the likelihood that a firm adopts the COLI shelter.

Structural Equivalence

In contrast to cohesion models of imitation, structural equivalence models suggest that firms imitate other organizations that have similar relationships with their environment, even in the absence of direct contacts (Galaskiewicz and Burt 1991). Structurally equivalent firms show similar adoption patterns because, in an attempt to maintain their competitive positions, firms

[14] See *AEP*, 136 F.Supp. 2d 762; *In re CM Holdings, Inc.*, 254 B.R. 578; *Winn-Dixie*, 113 T. C. 254 (1999); *Dow Chemical*, 250 F.Supp. 2d 748.

[15] AEP's CFO consulted with the Chairman and CEO of the firm before adopting the COLI program and appeared before the board of directors to explain program once it had been adopted (*AEP, Inc. v. United States*, 136 F.Supp. 2d 762).

[16] When evaluating whether to proceed with a broad-based, leveraged COLI plan, Camelot consulted with Ernst & Young (*In re CM Holdings, Inc.*, 254 B.R. 578).

imitate their successful peers (Burt 1987). Anecdotal evidence suggests that firms face significant pressure to imitate the tax savings strategies of their peers (Novack and Saunders 1998). Table 1 identifies 27 firms that recognized COLI tax savings as a rate reconciliation line-item in the 10-K tax footnotes, suggesting that even though tax shelters are typically characterized by secrecy, knowledge of COLI and its effect on ETR were reasonably public knowledge. Therefore, using industry membership as a proxy for structural equivalence (Fligstein 1985, 1990), I hypothesize:

H3: The presence of prior COLI shelter adopters in a firm's industry will increase the likelihood that a firm adopts the COLI shelter.

Spatial Proximity

Diffusion studies commonly find that spatially proximate actors influence each other (Rogers 2003). For example, Davis and Greve (1997) find that golden parachutes diffused via local business communities, and Burns and Wholey (1993) find regional differences in the adoption of matrix management.[17] Shared membership in a geographic region affords a high level of inter-firm interaction, resulting in the adoption by one firm in an area spurring others to do the same (Strang and Soule 1998). Using regional classifications developed by the Department of Commerce's Bureau of Economic Analysis (BEA region), I hypothesize:

H4a: The presence of prior COLI adopters in a firm's BEA region will increase the likelihood that a firm adopts the COLI shelter.

Regional differences in diffusion rates can also result from the geographic clustering of innovators. Local business elites are connected through a range of formal and informal institutions that facilitate communication, from the country club to local charity organizations. Consequently, those locales closest to an innovation's origin are likely sources of an innovation's earliest adopters (Rogers 2003). Three COLI court cases reveal the involvement of various insurance companies based in New York and New Jersey.[18] Therefore, I assume that the Mid-East BEA region is the center of COLI innovation and hypothesize:

H4b: The greater the geographic distance between the firm's headquarters and the Mid-East BEA region, the less likely the firm is to adopt the COLI shelter.

IV. MODELING COLI ADOPTION

Control Sample

To test whether imitation contributes to the spread of shelter use, I model a firm's decision to adopt COLI using a matched-control design. The foremost factors affecting whether a firm adopts or does not adopt a particular tax shelter are the presence of taxable income to shelter and compatibility between the technology of the shelter and the firm's organizational structure and asset and income mix.[19] The technology of COLI is based on maximizing the tax preferences related to insurance by investing in contracts on a large number of employees. To match COLI adopters to peer firms that could have benefited from adopting COLI but chose not to adopt, I form

---

[17] A parallel stream of literature in political science examines the regional diffusion of state policy innovations. See Berry and Berry (1990) for a review of this literature.

[18] Mutual Benefit (MBL) of New Jersey underwrote the policies purchased by Camelot Music and AEP, which was also approached with a deal funded by New York Life. AIG of New York underwrote the policies purchased by Winn-Dixie.

[19] Industry is one possible way to control for differences in shelter use arising from differences in the match between firms' operations and the technology of a specific shelter; however, to test whether COLI use spread along industry lines, I do not match COLI and control firms based on industry.

a control sample based on profitability and workforce size because these two firm characteristics are likely prerequisites for COLI adoption. First, I rank all Compustat firms into 20 groups based on *ROA* and *lnEMP* separately, where *ROA* is pre-tax book income minus minority interest divided by total assets ((PI − MII)/AT), and *lnEMP* is the log of the number of employees (EMP).[20] Then, I identify the *ROA* and *lnEMP* ranks of each COLI firm in year $t-1$ and include any firms with matching *ROA* and *lnEMP* ranks in that year in the control sample. The resulting one-to-many control sample includes 41 COLI firms and 330 control firms.[21]

To make the collection of board interlock data manageable, I develop a one-to-one control sample from the one-to-many sample and use this for my primary tests. For the one-to-one control sample, I choose the control firm with proxy statement data available on Lexis-Nexis that most closely matches each COLI firm's *ROA* and *lnEMP* in $t-1$. Given the difficulty in identifying shelter users, my control sample likely includes firms that used COLI, but whose sheltering activities have not been detected or disclosed.[22]

Cross-Sectional Model of Firms' Decisions to Adopt COLI

Initially, I employ the following cross-sectional logistic model to test whether social environment factors affect the likelihood of COLI adoption (H1–H4):

$$\begin{aligned} Log(P(t)/[1-P(t)]) = {} & \beta_0 + \beta_1 FOREIGN_{i,t-1} + \beta_2 RD_{i,t-1} + \beta_3 INDRANK_{i,t} + \beta_4 SGRWTH_{i,t} \\ & + \beta_5 ETR_{i,t-1} + \beta_6 BODLINK1_{i,t} + \beta_7 AUDITLINK1_{i,t} + \beta_8 INDLINK1_{i,t} \\ & + \beta_9 BEALINK1_{i,t} + \beta_{10} DISTANCE_{i,t} + \varepsilon_{i,t}, \end{aligned} \tag{2}$$

where $P(t)$ is the probability of adopting the COLI shelter.

Independent Variables

I test H1 by including *BODLINK1*, an indicator variable equal to 1 if the firm shared a board member with a COLI adopter at anytime between 1986 and 1995. I hand-collect information on each COLI and control firm's board of director membership from proxy statements. I assign each director a unique identifier and determine all board interlocks among the combined set of COLI and control firms. Both COLI-to-COLI interlocks and COLI-to-control interlocks result in *BODLINK1* equal to 1, while control-to-control interlocks do not.

BODLINK1 captures network ties through board interlocks generally. Here, I assume that the social ties and shared knowledge generated through board interlocks are not limited to actual interlock years, but spillover to surrounding years. In a later model specification, I construct a more restrictive measure of the influence of board interlocks that accounts for both the timing of shared board members and the timing of interlocked firms' COLI adoptions. Hypothesis 1 predicts that COLI adoptions spread via interlocking directorates; therefore, I expect the coefficient on *BODLINK1* to be positive.

I examine H2, the impact of indirect ties between prior and potential adopters via auditors, by including *AUDITLINK1*, an indicator variable equal to 1 if the firm's auditor in year *t* audited a prior COLI adopter. I obtain each firm's auditor during the year from Compustat's Company

---

[20] I require that firms are incorporated in the U.S. and have total assets greater than $1 million. These requirements apply to references to "all Compustat firms."

[21] Two of the original 43 COLI firms do not have sufficient data to construct all the explanatory variables included in the regression models.

[22] As a precaution, I exclude 19 firms that appear in my initial 213 COLI sample from the control sample even though I cannot confirm that these firms used broad-based, leveraged COLI. Inclusion of COLI firms in the control samples increases the noise in the data and would bias against finding results.

Auditor data set, data code AU. Consistent with H2, I expect a positive coefficient on *AUDITLINK1*.

To test H3, I include *INDLINK1*, an indicator variable equal to 1 if, at time *t*, a prior COLI adopter exists within the firm's industry. Industries are classified according to Barth et al. (1998). *INDLINK1* varies by firm and by year but does not vary across industry within a given year. Hypothesis 3 predicts that firms imitate their structurally equivalent peers in order to stay competitive; thus, I expect a positive coefficient on *INDLINK1*.

Similar to *INDLINK1*, I test H4a by including *BEALINK1*, an indicator variable equal to 1 if, at time *t*, a prior COLI adopter exists within the firm's BEA region. Hypothesis 4a posits that firms imitate their geographic neighbors; therefore, I expect a positive coefficient on *BEALINK1*. Hypothesis 4b predicts that shelter use will spread geographically starting from the source of the innovation. To test H4b, I construct *DISTANCE*, an ordinal variable equal to 1 for firms headquartered in the Mid-East BEA region; 2 for firms headquartered in the neighboring New England, Southeast, and Great Lakes BEA regions; 3 for firms headquartered in the Plains and Southwest BEA regions; and 4 for firms headquartered in the Rocky Mountain and Far West BEA regions. I expect the coefficient on *DISTANCE* to be negative.

Control Variables

The main control variables in Equation (2), *FOREIGN*, *RD*, *ETR*, and *SGRWTH*, are drawn from two recent studies that explicitly model firm characteristics associated with tax shelter participation (Wilson 2009; Lisowsky 2010). Both studies focus on factors that can be used to infer shelter use rather than factors related to the decision to adopt a shelter, and both studies compare non-shelter firms to a group of shelter firms, wherein the participants of several different types of tax shelters are combined together and treated as one class. One issue with grouping participants from different types of shelters is that each shelter has its own technology, and only firms whose traits match the technology of a particular shelter can use it. By modeling a firm's decision to adopt a single type of shelter, I hold differences in the technologies of various shelters constant. Consequently, I do not necessarily expect the same relationship between the control variables and the decision to adopt the COLI shelter that Wilson (2009) and Lisowsky (2010) hypothesize for shelter participation in general.[23]

Wilson (2009) finds that large book-tax differences (BTDs) are a marker of shelter use, and Lisowsky (2010) finds that tax shelter use is positively related to lagged ETR. Shelters that produce permanent book-tax differences are the most sought after (U.S. Treasury 1999; Weisbach 2002) as these reduce the firm's ETR and enable managers to simultaneously report low taxable income and high financial income. In terms of its financial statement effect, the COLI shelter is an archetype of recent corporate tax shelters.[24] I include the firm's prior year ETR, ETR_{t-1}, as a control variable.[25] However, since firms can adopt COLI either to obtain a competitive ETR or to maintain an already low and competitive ETR when an alternative tax-savings strategy dries up, I make no directional prediction for the relationship between ETR_{t-1} and the likelihood of adopting COLI.

Both Wilson (2009) and Lisowsky (2010) include ROA to capture profitability. Firms with greater profitability have more income to shelter. Since matching on ROA precludes including it as a control variable, I include growth in sales, *SGRWTH*, as an alternative measure of management's

[23] I omit several of the variables used by Wilson (2009) and Lisowsky (2010) because these variables capture the effect of tax shelter use rather than factors affecting the decision to adopt a shelter.

[24] See the Appendix for details on the financial statement reporting of COLI transactions.

[25] *ETR* is calculated as total income taxes (TXT) divided by pre-tax book income (PI). Following Gupta and Newberry (1997), I set *ETR* equal to 1 if TXT/PI is greater than 100 percent, and I set *ETR* equal to 0 when TXT/PI is negative.

incentive to shelter. I expect firms with higher sales growth to be more likely to adopt a shelter; therefore, I expect a positive relationship between *SGRWTH* and the likelihood of adopting COLI.

Since several corporate tax shelters involve foreign operations and intangibles, Wilson (2009) and Lisowsky (2010) predict a positive relationship between measures of these firm characteristics and tax shelter participation. However, compared to other codified and shelter-related, tax-savings strategies involving specific types of assets, liabilities, income, and expenses, COLI can be used by a broad set of firms. Therefore, I include two variables that proxy for firms' non-COLI tax savings opportunities, *RD* and *FOREIGN*, but I expect a negative, rather than a positive relationship between these firm characteristics and COLI adoption. To proxy for R&D activity, I include an indicator variable, RD_{t-1}, equal to 1 if the ratio of R&D expense to sales in the prior year is positive, and 0 otherwise. I proxy for the extent of a firm's foreign operations with an indicator variable, $FOREIGN_{t-1}$, equal to 1 if the ratio of the absolute value of a firm's foreign pretax income (PIFO) to the absolute value of its total pretax book income (PI – MII) is greater than 10 percent.

Finally, I include one additional variable suggested by diffusion theory, *INDRANK*. Larger organizations are generally more innovative (Rogers 2003). Furthermore, studies find that firms tend to imitate their successful and prestigious peers (Burns and Wholey 1993; Haveman 1993; Han 1994). To capture prestige, I rank all Compustat firms into 20 groups by year, industry, and total assets. I expect larger firms to adopt earlier and smaller firms to wait and adopt later, if at all; thus, I expect a positive coefficient on *INDRANK*.

Pattern of Board Interlocks

Figure 2 depicts all board interlocks within the combined COLI and one-to-one control sample during the time period 1986–1995. Of the 82 sample firms, 43 (52.4 percent) share a board member with another firm in the sample. COLI firms are shown as circles, including the year of COLI adoption, and control firms are shown as squares. The proportion of COLI firms with board ties to another sample firm (either COLI or control) does not significantly differ from the proportion of control firms with board ties to another sample firm (58.5 percent versus 46.3 percent; $z = 1.11$; $p = 0.27$). However, relative to control firms, COLI firms have significantly more interlocks to other COLI firms (48.78 percent versus 17.07 percent; $z = 3.05$; $p < 0.01$). The pattern of board interlocks depicted is consistent with H1, and suggests that COLI adopters are tied to one another via director contacts.

Pattern of COLI Adoption by Auditor, Industry, and Region

Table 2 shows the distribution of COLI adoptions across years, auditors, industries, and BEA regions. Notably, the first four COLI adopters in the sample each had different external auditors and, by 1990, COLI use had spread to clients of all of the former Big 6 auditors as well as BDO Seidman. The first four adopters are also in three different industries, and while over half of new COLI adoptions occur in the last three years, those adoptions only result in the spread of COLI use to two new industries. The pattern of COLI adoptions does not reveal an obvious concentration of early COLI adopters in any particular industry, auditor, or geographic region.

Table 3 provides further detail on the pattern of COLI use by comparing the distribution of COLI versus control firms across auditors, industries, and BEA geographic regions. Panel A shows that the breakdown of firms by auditor is very similar for the COLI and one-to-one control samples. Furthermore, the Big 6 auditors audited more than 95 percent of each of the three samples: COLI, one-to-one control, and one-to-many control. Panel B indicates that the proportion of firms in the textiles and chemicals industries is higher among COLI firms than for either of the

FIGURE 2
Diagram of Board Interlocks

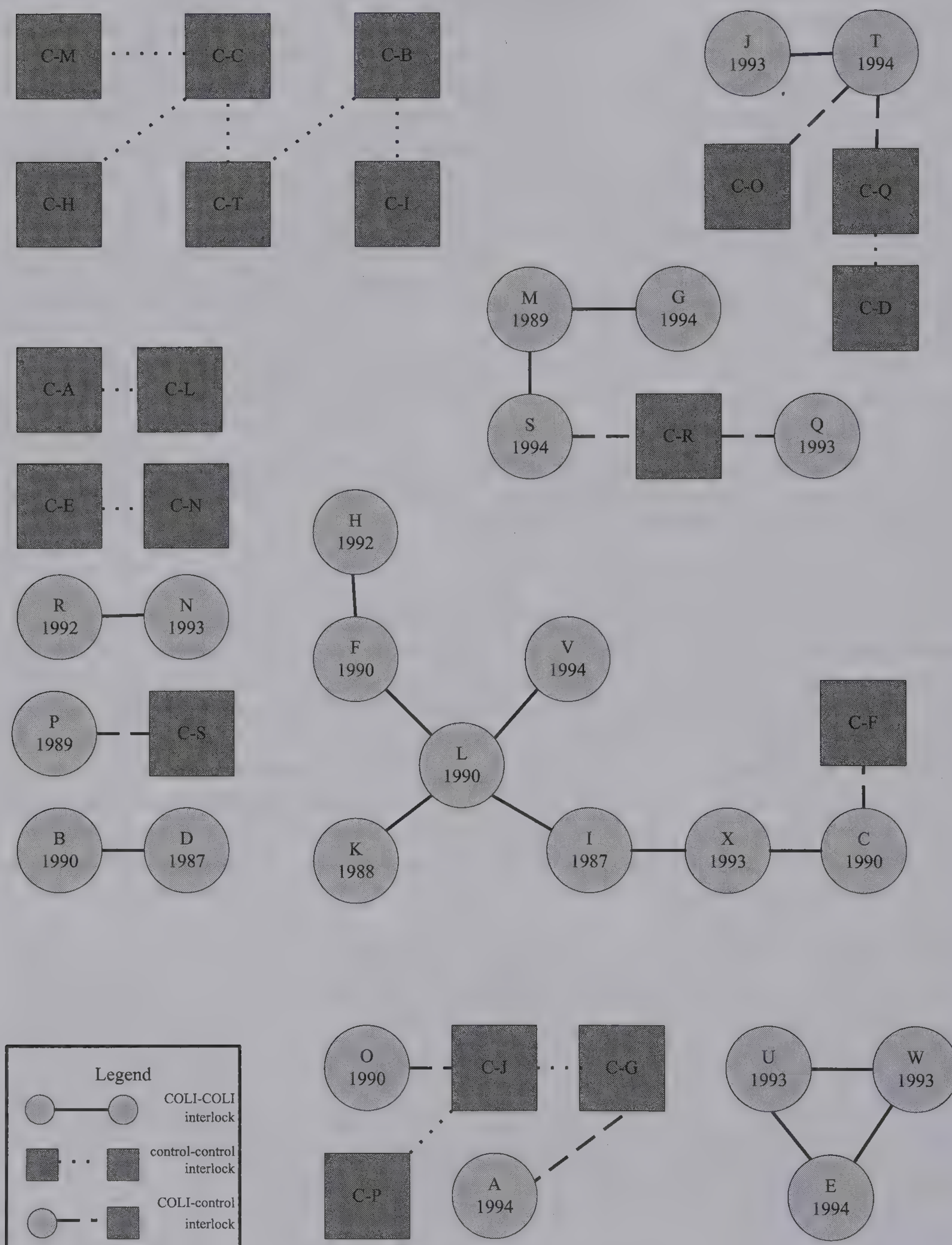


Figure 2 depicts all board interlocks within the combined COLI and control sample (n = 82 firms) during the time period 1986–1995. The one-to-one control sample is constructed by matching COLI firms and control firms based on *ROA* and *lnEMP* in the year prior to adoption ($t-1$) and proxy statement data availability.

TABLE 2

Pattern of New COLI Adoptions across Time, Auditors, Industries, and BEA Regions

	(1)	(2)	(3)	(4)
Year	**New COLI adopters**	**New Auditors represented**	**New Industries represented**	**New BEA Regions represented**
1986	1	1	1	1
1987	2	2	1	2
1988	1	1	1	0
1989	5	2	3	2
1990	7	1	2	0
1991	1	0	1	0
1992	2	0	0	1
1993	11	1	2	1
1994	9	0	0	0
1995	2	0	0	0
Total	41	8	11	7

Sample includes 41 COLI adopters with data to construct all explanatory variables. Column (1) shows the number of new COLI adopters by year. Broad-based, leveraged COLI was developed in 1985 and 1986 in response to restrictions enacted in TRA86 and was effectively shut down by further tax law changes in 1996 (HIPAA). Columns (2)–(4) show the number of new auditors, industries, and BEA regions represented by those adopters. Auditor is determined from Compustat data code AU. Industry classifications are based on Barth et al. (1998). BEA region classifications are obtained from the U.S. Department of Commerce's Bureau of Economic Analysis and assigned based on Compustat data code STATE.

two control samples, but the distribution of firms across industries is otherwise similar among the three groups. Contrary to H2 and H3, the descriptive data in Table 2 and Table 3 show no particular concentration of COLI adopters by auditor or industry.

Panel C of Table 3 shows some difference between the COLI and control samples in the distribution of firms across geographic region, consistent with H4a and H4b. Notably, although 10–12 percent of the control firms are headquartered in the Far West region, there are no COLI firms headquartered there. Also, relative to the one-to-one control sample, there are significantly more COLI firms headquartered in the Great Lakes region ($\chi^2 = 3.36$; p = 0.07).

Descriptive Statistics

Table 4 provides descriptive statistics for both the COLI and one-to-one control samples, measured in year t. There is no statistical difference between the COLI and control samples in mean lagged profitability (ROA_{t-1}) or mean lagged workforce size ($lnEMP_{t-1}$), suggesting an effective matching procedure. Furthermore, there is no statistical difference in the average size of the COLI and control firms. Inconsistent with my expectations, control firms have significantly higher sales growth (*SGRWTH*) relative to COLI firms (12.5 percent versus 7.6 percent; two-tailed p = 0.04). Given that the COLI and control firms have similar profitability, one explanation for the lower sales growth among COLI firms is that COLI firms, facing lower growth prospects, focus on maintaining profitability by reducing expenses, specifically reducing tax expense through the use of a tax shelter.

Consistent with H1, tests for differences in proportions indicate that COLI-to-COLI board interlocks are significantly more common than COLI-to-control board interlocks. Contrary to H2, but consistent with the distribution of observations across auditors reported in Table 3, the proportion of COLI firms with links to a prior COLI adopter via shared auditor is actually lower than

TABLE 3

Distribution of COLI and Control Firms across Auditor, Industries, and BEA Regions[a]

Panel A: Distribution of Observations across Audit Firms (year *t*)

	COLI		One-to-One Control Sample		One-to-Many Control Sample	
Auditor[b]	**Obs.**	**Percent of Sample**	**Obs.**	**Percent of Sample**	**Obs.**	**Percent of Sample**
Arthur Andersen	5	12.20%	7	17.07%	66	20.00%
Arthur Young	0	0.00%	0	0.00%	2	0.61%
Coopers & Lybrand	5	12.20%	2	4.88%	39	11.82%
Ernst & Young	13	31.69%	12	29.27%	59	17.88%
Deloitte & Touche	4	9.76%	4	9.76%	55	16.67%
KPMG	7	17.07%	8	19.51%	43	13.03%
PricewaterhouseCoopers	5	12.20%	6	14.63%	50	15.15%
Touche Ross	0	0.00%	1	2.44%	4	1.21%
Other	0	0.00%	0	0.00%	8	2.42%
BDO Seidman	1	2.44%	0	0.00%	0	0.00%
Grant Thornton	0	0.00%	0	0.00%	2	0.61%
Kenneth Leventhal	1	2.44%	0	0.00%	1	0.30%
Laventhol and Horwath	0	0.00%	1	2.44%	1	0.30%
Total	41	100.00%	41	100.00%	330	100.00%

(continued on next page)

Panel B: Distribution of Observations across Industries

Industry Description[c]	Primary SIC Codes	COLI Obs.	COLI Percent of Sample	One-to-One Control Sample Obs.	One-to-One Control Sample Percent of Sample	One-to-Many Control Sample Obs.	One-to-Many Control Sample Percent of Sample
Mining and construction	1000–1299, 1400–1999	1	2.44%	2	4.88%	13	3.94%
Food	2000–2111	1	2.44%	2	4.88%	21	6.36%
Textiles, printing and publishing	2200–2780	7	17.07%	2	4.88%	29	8.79%
Chemicals	2800–2824, 2840–2899	5	12.20%	2	4.88%	11	3.33%
Pharmaceuticals	2830–2836	0	0.00%	0	0.00%	8	2.42%
Extractive industries	2900–2999, 1300–1399	1	2.44%	1	2.44%	7	2.12%
Durable manufacturers	3000–3569, 3580–3669, 3680–3999	10	24.38%	12	29.25%	78	23.64%
Transportation	4000–4899	2	4.88%	2	4.88%	16	4.85%
Utilities	4900–4999	2	4.88%	0	0.00%	16	4.85%
Retail	5000–5999	3	7.32%	4	9.76%	29	8.79%
Computers	7370–7379, 3570–3579, 3670–3679	7	17.07%	10	24.39%	72	21.82%
Services	7000–7369, 7380–8999	2	4.88%	4	9.76%	30	9.09%
Other	> 9000	0	0.00%	0	0.00%	0	0.00%
Total		41	100.00%	41	100.00%	330	100.00%

(continued on next page)

Panel C: Distribution of Observations across BEA Regions

	COLI		One-to-One Control Sample		One-to-Many Control Sample	
BEA Region[d]	**Obs.**	**Percent of Sample**	**Obs.**	**Percent of Sample**	**Obs.**	**Percent of Sample**
New England Region	4	9.76%	6	14.63%	32	9.70%
Mideast Region	10	24.39%	11	26.83%	65	19.70%
Great Lakes Region	13	31.70%	6	14.63%	70	21.19%
Plains Region	3	7.32%	1	2.44%	25	7.58%
Southeast Region	9	21.95%	7	17.07%	65	19.70%
Southwest Region	1	2.44%	3	7.32%	29	8.79%
Rocky Mountain Region	1	2.44%	2	4.88%	9	2.73%
Far West Region	0	0.00%	5	12.20%	35	10.61%
Total	41	100.00%	41	100.00%	330	100.00%

[a] The one-to-many control sample is constructed by ranking all Compustat firms into 20 groups based on *ROA* and *lnEMP* separately, identifying the *ROA* and *lnEMP* ranks of each COLI firm in year t–1 and matching control firms based on those *ROA* and *lnEMP* ranks. *ROA* is pre-tax book income minus minority interest divided by total assets ((PI – MII)/AT). *lnEMP* is the log of the number of employees (EMP). The one-to-one control sample is constructed by choosing the closest *ROA* and *lnEMP* match with available proxy statements from the one-to-many control sample.

[b] Auditor is determined from Compustat code AU. Firm mergers and acquisitions are reflected above.

[c] Industry classifications are based on Barth et al. (1998).

[d] BEA region classifications are obtained from the U.S. Department of Commerce's Bureau of Economic Analysis and assigned based on Compustat data code STATE.

TABLE 4

Descriptive Statistics of Regression Variables (COLI and control firms, year t)

Panel A: Distribution of Continuous and Ordinal Variables

	COLI Sample (n = 41)				Control Sample (n = 41)			
Variable	**Mean**	**Q1**	**Median**	**Q3**	**Mean[a]**	**Q1**	**Median**	**Q3**
Total Assets ($billions)	3.020	0.534	1.141	3.096	1.900	0.324	1.075	2.455
$LnEMP_{t-1}$	2.543	1.581	2.303	3.384	2.408	1.478	2.485	3.045
ROA_{t-1}	0.103	0.064	0.101	0.138	0.100	0.065	0.101	0.130
ETR_{t-1}								
SGRWTH	0.076	0.018	0.061	0.113	0.125**	0.052	0.090	0.170
INDRANK	8.976	9.000	9.000	10.000	8.780	8.000	9.000	10.000
DISTANCE	1.902	2.000	2.000	2.000	2.171	1.000	2.000	3.000

Panel B: Frequency of Indicator Variables

Variable	COLI Sample (n = 41)	Control Sample (n = 41)	Difference in Proportions	z-statistic[b]
$FOREIGN_{t-1}$	26.83%	41.46%	−14.63%	−1.40
RD_{t-1}	48.78%	39.02%	9.76%	0.89
BODLINK	48.78%	17.07%	31.71%	3.05***
AUDITLINK	80.49%	92.68%	−12.20%	−1.62
INDLINK	65.85%	75.61%	−9.76%	−0.92
BEALINK	80.49%	70.73%	9.76%	1.03

[a] *, **, *** Indicate significance at the 0.10, 0.05 and 0.01 levels for two-tailed, t-tests of difference in the means.

[b] *, **, *** Indicate significance at the 0.10, 0.05 and 0.01 levels for two-tailed, tests of differences in proportions. The percentages reported represent the proportion of observations in each subsample where the indicator variable equals 1.

Variable Definitions

$LnEMP_{t-1}$ = lagged value of the log of the number of employees (EMP);
ROA_{t-1} = lagged value of pre-tax book income (PI) minus minority interest (MII), divided by the average total assets (AT);
ETR_{t-1} = lagged value of total income taxes (TXT) divided by pre-tax book income (PI);

(continued on next page)

SGRWTH = sales (SALE) in year t minus sales in year $t-1$, divided by sales in year $t-1$;
INDRANK = decile-ranking resulting from ranking all Compustat firms with assets greater than $1 million by year, industry, and total assets (AT);
DISTANCE = ordinal variable equal to 1 for firms headquartered in the Mid-East BEA region, 2 for firms headquartered in the neighboring New England, Southeast, and Great Lakes BEA regions, 3 for firms headquartered in the Plains and Southwest BEA regions, and 4 for firms headquartered in the Rocky Mountain and Far West BEA regions;
$FOREIGN_{t-1}$ = indicator variable equal to 1 if the absolute value of foreign pre-tax income (PIFO) divided by the absolute value of total pre-tax income (PI) is greater than 10 percent;
RD_{t-1} = indicator variable equal to 1 if R&D expense (XRD) is positive;
BODLINK = indicator variable equal to 1 if the firm has a board interlock with a COLI firm at any time during the period from 1986–1995;
AUDITLINK = indicator variable equal to 1 if the firm's auditor in year t has audited a prior COLI adopter;
INDLINK = indicator variable equal to 1 if, in year t, there is a prior COLI adopter in the firm's industry; and
BEALINK = indicator variable equal to 1 if, in year t, there is a prior COLI adopter in the firm's BEA region.

the proportion of control firms with such links (80.49 percent versus 92.68 percent; $z = -1.62$; two-tailed $p = 0.11$). There is no significant difference between the COLI and control sample in the proportion of firms with links to prior COLI adopters via industry or BEA region.

Cross-Sectional Logit Results

Table 5 reports results for estimating a series of nested cross-sectional logistic regressions constructed from Equation (2). All models have reasonable explanatory power with pseudo-R^2 values ranging from 10.45 percent to 16.33 percent.[26] Of the firm-specific characteristics, only $FOREIGN_{t-1}$ is significant in all models. As expected, COLI adopters are less likely to have significant foreign operations relative to non-adopters. Wilson (2009) and Lisowsky (2010) generally find that foreign income is positively associated with shelter use; however, their shelter samples group a variety of shelters together, including some that can only be used by firms with foreign assets or income. As such, the positive association they find between foreign operations and tax shelter use may reflect the technology match between certain shelters and certain firms rather than a determinant of tax shelter adoption generally.

The results for RD_{t-1} and *INDRANK* are not in the predicted direction and vary across the models. Wilson (2009) also reports mixed results for the effect of R&D intensity on tax shelter use. *INDRANK* is based on the firm's size (AT) relative to the other firms in its industry. Table 3 and Table 4 show that COLI and control firms are distributed evenly across industries and are not statistically different in size. Furthermore, I match firms based on *ROA* and *lnEMP*, diminishing differences in *INDRANK* across the two groups. The coefficient on ETR_{t-1} is not significant in any of the models. In untabulated results, I substitute ETR_{t-1} with measures of the firm's lagged ETR adjusted for the lagged industry mean (median) ETR, and those measures are also not significant. Since I match firms based on profitability, which is a major determinant of ETR, there is likely little systematic variation in the ETR between the control and COLI firms.[27]

Consistent with H1, which predicts that board interlocks impact firms' COLI adoption decisions, the coefficient on *BODLINK1* is positive and significant (1.02; $z = 1.72$; one-tailed $p = 0.04$). Holding all indicator variables equal to 1 and all other variables at their median values, a change in *BODLINK1* from 0 to 1 is associated with a 73.6 percent increase (from 33.7 percent to 58.5 percent) in the probability that a firm adopts the COLI shelter. *BODLINK1* is a broad measure that generally proxies for the set of social ties between firms. In the event history analysis presented below, I explore the association between board interlocks and the decision to adopt the COLI shelter further.

Hypothesis 2 predicts that auditors impact firms' COLI adoption decisions. However, the coefficient on *AUDITLINK1* is negative rather than the predicted positive association; suggesting that prior and potential COLI adopters are not linked via shared auditors. Although the results in Table 5 cannot be generalized to other tax shelters, individual or corporate, COLI use does not appear to have spread through any particular public accounting firm.

Hypothesis 3 predicts the likelihood of adopting COLI increases when there is a prior adopter in a firm's industry. A positive coefficient on *INDLINK1* would imply that firms imitate their structurally equivalent peers' tax sheltering activities. However, as Table 5 shows, the coefficient on *INDLINK1* is not statistically significant. One possible explanation is that industry does not accurately capture a firm's structurally equivalent peer group. Another is that, among potential diffusing mechanisms, the spread of COLI is better characterized by a direct-contact cohesion

---

[26] For the least restrictive model, including *BODLINK1* and *DISTANCE*, the Hosmer and Lemeshow goodness-of-fit χ^2 equals 76.4, with a p-value for the null hypothesis equal to 0.40, indicating that the model is reasonably well fitted.

[27] Both COLI and control firms may be engaged in other, non-COLI aggressive tax savings strategies either before, during, or after the COLI adoption decision; these activities could also impact ETR.

TABLE 5

Cross-Sectional Logit Model of COLI Adoption

$$\mathrm{Log}(P(t)/[1-P(t)]) = \beta_0 + \beta_1 FOREIGN_{i,t-1} + \beta_2 RD_{i,t-1} + \beta_3 INDRANK_{i,t} + \beta_4 SGRWTH_{i,t} + \beta_5 ETR_{i,t-1} + \beta_6 BODLINK1_{i,t} + \beta_7 AUDITLINK1_{i,t} + \beta_8 INDLINK1_{i,t} + \beta_9 BEALINK1_{i,t} + \beta_{10} DISTANCE_{i,t} + \varepsilon_{i,t}$$

Variable	Pred. Sign	Firm Characteristics		H1: BOD Interlock		H2: Auditor		H3: Industry	
		Coeff	z-stat	Coeff	z-stat	Coeff	z-stat	Coeff	z-stat
Intercept		−1.958	−1.08	−1.573	−0.85	−1.466	−0.77	−1.980	−1.08
$FOREIGN_{t-1}$	−	−1.489	−2.36***	−1.205	−1.84**	−1.603	−2.43***	−1.466	−2.32***
RD_{t-1}	−	0.848	1.34	0.585	0.89	0.779	1.20	0.900	1.41
INDRANK	+	0.252	1.20	0.160	0.74	0.398	1.71**	0.303	1.39*
SGRWTH	+	−5.655	−2.08	−4.000	−1.43	−6.672	−2.32	−5.600	−2.08
ETR_{t-1}	?	1.133	0.49	1.044	0.44	0.750	0.31	1.163	0.50
BODLINK1	+			1.020	1.72**				
AUDITLINK1	+					−1.673	−1.95		
INDLINK1	+							−0.670	−1.19
BEALINK1	+								
DISTANCE	−								
Goodness of Fit Statistics									
Pseudo-R^2		10.45%		13.14%		14.24%		11.73%	
Chi-square test		11.88**		14.94**		16.19**		13.33**	
Area under ROC curve		0.70		0.73		0.72		0.71	

Variable	Pred. Sign	H4a: BEA Region		H4b: Distance		H1 and H4b	
		Coeff	z-stat	Coeff	z-stat	Coeff	z-stat
Intercept		−2.724	−1.43	−0.693	−0.35	−0.170	−0.08
$FOREIGN_{t-1}$	−	−1.615	−2.48***	−1.631	−2.51***	−1.344	−2.00***

(continued on next page)

Variable	Pred. Sign	H4a: BEA Region Coeff	H4a: BEA Region z-stat	H4b: Distance Coeff	H4b: Distance z-stat	H1 and H4b Coeff	H1 and H4b z-stat
RD_{t-1}	−	1.113	1.66*	0.883	1.37	0.621	0.92
INDRANK	+	0.242	1.14	0.259	1.19	0.159	0.70
SGRWTH	+	−6.072	−2.08	−6.531	−2.20	−4.687	−1.53
ETR_{t-1}	?	1.441	0.62	0.733	0.31	0.593	0.25
BODLINK1	+					1.110	1.83**
AUDITLINK1	+						
INDLINK1	+						
BEALINK1	+	0.931	1.54*				
DISTANCE	−			−0.529	−1.74**	−0.577	−1.82**
Goodness of Fit Statistics							
Pseudo-R^2		12.62%		13.30%		16.33%	
Chi-square test		14.35**		15.11**		18.56***	
Area under ROC curve		0.70		0.73		0.74	

*, **, *** Indicate significance at the 0.10, 0.05, and 0.01 levels in two-tailed tests (one-tailed when predicted).

The sample includes 41 COLI firms and 41 control firms from a sample matched on ROA and lnEMP in year $t-1$ (the year before adoption).

Variable Definitions:

$FOREIGN_{t-1}$ = indicator variable equal to 1 if the absolute value of foreign pre-tax income (PIFO) divided by the absolute value of total pre-tax income (PI) is greater than 10 percent;
RD_{t-1} = indicator variable equal to 1 if R&D expense (XRD) is positive;
INDRANK = decile-ranking resulting from ranking all Compustat firms by year, industry, and total assets (AT);
SGRWTH = sales (SALE) in year t minus sales in year $t-1$, divided by sales in year $t-1$;
ETR_{t-1} = lagged value of total income taxes (TXT) divided by pre-tax book income (PI);
BODLINK1 = indicator variable equal to 1 if the firm has a board interlock with a COLI firm at any time during the period from 1986–1995;
AUDITLINK1 = indicator variable equal to 1 if the firm's auditor in year t has audited a prior COLI adopter;
INDLINK1 = indicator variable equal to 1 if, in year t, there is a prior COLI adopter in the firm's industry;
BEALINK1 = indicator variable equal to 1 if, in year t, there is a prior COLI adopter in the firm's BEA region; and
DISTANCE = ordinal variable equal to 1 for firms headquartered in the Mid-East BEA region, 2 for firms headquartered in the neighboring New England, Southeast, and Great Lakes BEA regions, 3 for firms headquartered in the Plains and Southwest BEA regions, and 4 for firms headquartered in the Rocky Mountain and Far West BEA regions.

model than by an indirect structural equivalence model. Similar to H3, H4a predicts the likelihood of adopting COLI increases when there is a prior adopter in a firm's geographic region. The coefficient on *BEALINK1* is positive, but not significant.

Hypothesis 4b predicts that exposure to COLI through local social networks will be strongest in locales near the large insurance firms underwriting COLI policies; thus, firms with headquarters farther away from the Mid-East BEA region will be less likely to adopt the COLI shelter. Consistent with H4b, the coefficient on *DISTANCE* is negative and significant (–0.529; $z = -1.74$; one-tailed $p = 0.04$). The last column in Panel B of Table 5 shows that *BODLINK1* and *DISTANCE* remain statistically significant when they are included in the model together, indicating that COLI adoptions are influenced by two complementary diffusing mechanisms. Holding all indicator variables equal to 1 and all other variables, including *DISTANCE*, at their median values, a change in *BODLINK1* from 0 to 1 is associated with a 84 percent increase (from 31.9 percent to 58.7 percent) in the probability that a firm adopts the COLI shelter. Holding all indicator variables, including *BODLINK1*, equal to 1 and all other variables at their median values, a one-unit change in *DISTANCE*, from *DISTANCE* = 2 to *DISTANCE* = 3, is associated with a 24.5 percent decrease (from 58.7 percent to 44.3 percent) in the probability that a firm adopts the COLI shelter.

The evidence in Table 5 suggests that COLI adoption spread among firms with shared board members and spread out geographically from the Mid-East region. Overall, results from using a one-to-one matched control sample and estimating cross-sectional models based on Equation (2) are consistent with H1 and H4b, but are not consistent with H2, H3, and H4a. In untabulated tests, I use a one-to-many control sample to test H2, H3, H4a, and H4b. The coefficient on *BEALINK1* is positive and significant (0.79; $z = 1.86$; one-tailed $p = 0.03$), and the coefficient on *DISTANCE* is negative and significant (-0.50; $z = -2.33$; one-tailed $p = 0.01$). The results using a one-to-many control sample are consistent with both H4a and H4b, but are not consistent with H2 and H3.

V. EVENT HISTORY ANALYSIS

Model Specification

The cross-sectional model presented above tests whether certain social environment factors affect the likelihood of adopting COLI at time *t*, but does not test how those factors affect the timing of COLI adoptions. Below, I follow most diffusion studies and employ an event history or duration model. The variable of interest in a duration model is the length of time, *T*, that elapses between the time a firm is first "at risk" of adopting a new practice and the time, *t*, at which the firm adopts the practice. Duration data can be used to estimate a hazard rate, $h(t)$, the rate at which the duration ends in the interval $[t, t+\Delta]$, given that the duration has not ended prior to the beginning of this interval:

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t + \Delta t > T \geq t | T \geq t)}{\Delta t}. \tag{3}$$

I am interested in the hazard rate, the probability that a firm will adopt the COLI shelter at time *t*, given that it has not previously adopted the COLI shelter. Although event history analysis often utilizes continuous-time hazard-rate models, the data in this study are better suited to a discrete-time model because both the time of adoption and the related explanatory variables are observed at annual intervals only (Allison 1984).[28] I construct a discrete-time event history model in which each firm contributes an observation for each year it is "at risk" of adopting a COLI

[28] Discrete-time hazard models have been used to examine the diffusion of a variety of corporate practices (e.g., Fligstein 1985; Mezias 1990; Rao and Sivakumar 1999) as well as the adoption pattern of various state policies (e.g., Berry and

shelter. Court documents indicate that the COLI shelter was developed in response to TRA86, and I identify at least one firm that appears to adopt COLI in 1986. Therefore, I assume that firms are "at-risk" if they have not adopted the COLI shelter and t is greater than 1985 and estimate a logistic regression of the following form:

$$\log(P(t)/[1 - P(t)]) = \sum a(t) + \sum b_i x_i(t) + \sum c_j x_j, \tag{4}$$

where $P(t)$ is the probability of adopting the COLI shelter, b_i is the set of coefficients for explanatory variables $x_i(t)$ that do change with time, c_j is the set of coefficients for explanatory variables x_j that do not change over time, and $a(t)$ represents separate constants for each period t. Specifically, I estimate nested models derived from the following logistic regression:

$$\begin{aligned} Log(P(t)/\ [1 - P(t)])) = {} & \beta_0 + \beta_1 FOREIGN_{i,t-1} + \beta_2 RD_{i,t-1} + \beta_3 INDRANK_{i,t} + \beta_4 SGRWTH_{i,t} \\ & + \beta_5 ETR_{i,t-1} + \beta_6 - \beta_{14} yr\ indicators + \beta_{15} BODLINK2_{i,t} \\ & + \beta_{16} AUDITLINK2_{i,t} + \beta_{17} INDLINK2_{i,t} + \beta_{18} BEALINK2_{i,t} \\ & + \beta_{19} DISTANCE_{i,t} + \varepsilon_{i,t}. \end{aligned} \tag{5}$$

The discrete hazard model specified in Equation (5) accommodates time-varying covariates by splitting the history of each sample firm into one-year records or spells, with all spells except the year of adoption coded as right-censored.[29]

Duration models commonly assume that the baseline hazard rate is time-invariant and that the rate of adoption varies with firm characteristics, but is otherwise constant across time. However, as Figure 1 indicates, COLI adoptions appear to be related to firms' expectations about legislative measures to restrict the preferential tax treatment afforded to life insurance. Therefore, I include year indicators to determine how the hazard function changes over time. If the threat of potential adverse tax law changes prompted firms to complete COLI programs in the hopes of a grandfathering provision, then the rate of COLI adoption should be higher in 1989 and 1990 relative to other years. If Congress' failure to pass legislation curbing COLI benefits in 1991 and 1992 reduced the uncertainty surrounding COLI, then the rate of COLI adoption should be higher in 1993 relative to other years.

Independent Variables

In the cross-sectional model, I test H1 using a broad measure of board interlock influence, *BODLINK1*, that does not vary across time. In the event history model, I employ a more restrictive measure of board interlock influence that accounts for the timing of board interlocks and COLI adoptions. *BODLINK2* is equal to the number of board interlocks the sample firm has in year t with firms that have adopted COLI prior to year t. Consistent with H1, I expect a positive coefficient on *BODLINK2*.

The cross-sectional model also employs dichotomous variables to test H2, H3, and H4a. In the event history model, I use three alternative measures to test these hypotheses: *AUDITLINK2*, *INDLINK2*, and *BEALINK2*. *AUDITLINK2* is equal to the number of prior COLI adopters audited by the firm's auditor in year t, scaled by the total number of Compustat firms audited by the firm's

Berry 1992, 1994; Mintrom and Vergari 1998). See Box-Steffensmeier and Jones (1997) for an in-depth discussion of the benefit of using discrete-time models for certain data and Omer and Shelley (2004) for an example of the use of a discrete-time specification in the accounting literature.

[29] COLI adopters contribute firm years in proportion to the timing of adoption; a firm that adopts in 1986 contributes one firm year, whereas a firm that adopts in 1995 contributes ten firm years, nine coded as 0 and one coded as 1. Non-adopters contribute ten firm years, the length of the window, all coded as 0.

auditor in year *t*. If COLI spreads via shared auditors (H2), then the rate of COLI adoption should be higher for firms whose auditors have prior contact with COLI, resulting in a positive coefficient for *AUDITLINK2*.

INDLINK2 is equal to the number of prior COLI adopters in a firm's industry in year *t*, scaled by the total number of Compustat firms in the firm's industry in year *t*. If firms imitate their structurally equivalent peers captured by shared industry membership (H3), then the rate of COLI adoption should be higher for firms in industries with a greater frequency of prior COLI adopters, resulting in a positive coefficient for *INDLINK2*. Similarly, *BEALINK2* equals the number of prior COLI adopters in the firm's BEA region in year *t*, scaled by the total number of Compustat firms in the firm's BEA region in year *t*. If firms imitate their geographically proximate peers (H4a), then the coefficient on *BEALINK2* will be positive.

As in the cross-sectional model, I include *DISTANCE*, an ordinal variable that proxies for the geographic distance between a firm's headquarters and the center of early COLI activity. I expect firms headquartered within or close to the BEA's Mid-East region to adopt earlier relative to firms headquartered further away (H4b); therefore, I expect the rate of COLI adoption to be negatively associated with *DISTANCE*.

Discrete-Time Hazard Logit Results

Table 6 reports results from estimating the discrete-time hazard model in Equation (5).[30] All models have reasonable explanatory power, with pseudo-R^2 values ranging from 15.62 percent to 19.29 percent. The coefficients on the year indicators for 1989, 1990, 1993, and 1994 are positive and significant, indicating that the hazard rate, the likelihood of adopting COLI in year *t* given that the firm has not adopted COLI prior to year *t*, is significantly greater in those years relative to the base year, 1986. This result is consistent with firms timing their adoptions in response to concerns over legislative restrictions. As in the cross-sectional model, $FOREIGN_{t-1}$ is the only control variable that is significant in the predicted direction across all models.

Consistent with H1, the coefficient on *BODLINK2* is positive and significant (1.67; $z = 3.72$; $p < 0.01$); the conditional likelihood of adopting COLI in year *t* is greater for firms with board ties to prior COLI adopters. In marginal effect terms, given that the firm has not already adopted COLI, in 1993, a firm with one board tie to a prior COLI adopter is 2.4 times as likely to adopt COLI relative to a firm with no board ties (increase in probability of adopting from 11.97 percent to 40.74 percent).[31]

Contrary to H2 and H3, but similar to the results in Table 5, the coefficients on *AUDITLINK2* and *INDLINK2* in the discrete-time hazard model are not significant. While *DISTANCE* is significant in the cross-sectional models using a one-to-one control sample and both *DISTANCE* and *BEALINK1* are significant in the cross-sectional model using a one-to-many control sample, neither *DISTANCE* nor *BEALINK2* is significant in the discrete-time hazard model. One interpretation of these mixed results is that *DISTANCE* affects the likelihood of adopting COLI but, contrary to expectation, *DISTANCE* is not related to the timing of adoptions.[32]

---

[30] Three COLI firms do not have data available to calculate the explanatory variables in all years from 1986 to 1996; thus, the sample used to estimate Equation (5) consists of 629 firm-year observations representing 38 COLI firms and 38 matched control firms.

[31] The marginal effect is calculated by setting the indicator variable for 1993, $FOREIGN_{t-1}$ and RD_{t-1} equal to 1, all other year indicator variables equal to 0, and all continuous and ordinal variables equal to their median values.

[32] In untabulated results using a Tobit specification, *DISTANCE* is negatively and significantly related to *DURATION*, which equals the total number of years from the time the firm started using COLI until 1996 for COLI firms, and 0 for control firms (coeff $= -1.88$; $t = -1.99$; one-tailed $p = 0.03$). The Tobit results are consistent with H4b, firms headquartered farther away from the Mid-East BEA region adopt later.

TABLE 6

Discrete-Time Hazard Model

$$\text{Log}(P(t)/[1 - P(t)]) = \beta_0 + \beta_1 FOREIGN_{i,t-1} + \beta_2 RD_{i,t-1} + \beta_3 INDRANK_{i,t} + \beta_4 SGRWTH_{i,t} + \beta_5 ETR_{i,t-1} + \beta_6 - \beta_{14} \textit{yr indicators} + \beta_{15} BODLINK2_{i,t} + \beta_{16} AUDITLINK2_{i,t} + \beta_{17} INDLINK2_{i,t} + \beta_{18} BEALINK2_{i,t} + \beta_{19} DISTANCE_{i,t} + \varepsilon_{i,t}.$$

Variable	Pred. Sign	Base		H1: BOD Interlock		H2: Auditor		H3: Industry	
		Coeff	z-stat	Coeff	z-stat	Coeff	z-stat	Coeff	z-stat
Intercept		−6.247	−3.07***	−5.991	−2.92***	−6.419	−3.14***	−6.258	−3.09***
$FOREIGN_{t-1}$	−	−1.204	−2.40***	−1.084	−2.04***	−1.155	−2.29***	−1.202	−2.39***
RD_{t-1}	−	0.766	1.92	0.730	1.71	0.770	1.91	0.745	1.69
INDRANK	+	0.176	1.06	0.134	0.82	0.191	1.15	0.178	1.08
SGRWTH	+	−2.044	−1.91	−1.304	−1.22	−1.975	−1.81	−2.041	−1.90
ETR_{t-1}	?	1.036	0.87	1.217	1.05	1.091	0.90	1.037	0.87
D-1989		1.980	1.71*	1.812	1.55	2.045	1.75*	1.974	1.70*
D-1990		2.131	1.89*	1.916	1.70*	2.263	1.97**	2.115	1.85*
D-1993		2.962	2.70***	2.796	2.54**	3.236	2.79***	2.928	2.57**
D-1994		3.046	2.73***	2.665	2.39**	3.454	2.76***	2.991	2.56**
BODLINK2	+			1.670	3.72***				
AUDITLINK2	+					−0.669	−0.81		
INDLINK2	+							0.058	0.19
BEALINK2	+								
DISTANCE	−								
Goodness of Fit Statistics									
Pseudo-R^2		15.62%		19.17%		15.86%		15.62%	
Chi-square test		35.60***		51.36***		36.50***		37.59***	
Area under ROC curve		0.80		0.81		0.80		0.80	

(continued on next page)

Variable	Pred. Sign	H4a: BEA Region Coeff	H4a: BEA Region z-stat	H4b: Distance Coeff	H4b: Distance z-stat	H1 and H4b Coeff	H1 and H4b z-stat
Intercept		−6.341	−3.10***	−5.852	−2.75***	−5.761	−2.67***
$FOREIGN_{t-1}$	−	−1.200	−2.44***	−1.186	−2.36***	−1.072	−2.00***
RD_{t-1}	−	0.711	1.74	0.711	1.76	0.695	1.59
INDRANK	+	0.186	1.11	0.195	1.19	0.145	0.90
SGRWTH	+	−2.096	−1.93	−2.014	−1.84	−1.297	−1.20
ETR_{t-1}	?	1.112	0.99	0.858	0.71	1.115	0.94
D-1989		1.914	1.64*	1.931	1.67*	1.791	1.53
D-1990		1.981	1.75*	2.094	1.86*	1.902	1.69*
D-1993		2.691	2.37**	2.919	2.66***	2.776	2.52**
D-1994		2.694	2.30**	3.027	2.73***	2.664	2.40**
BODLINK2	+					1.621	3.53***
AUDITLINK2	+						
INDLINK2	+						
BEALINK2	+	0.583	1.08				
DISTANCE	−			−0.224	−1.10	−0.129	−0.60
Goodness of Fit Statistics							
Pseudo-R^2		16.02%		16.02%		19.29%	
Chi-square test		38.05**		37.66***		49.86***	
Area under ROC curve		0.80		0.80		0.81	

*, **, *** Indicate significance at the 0.10, 0.05, and 0.01 levels in two-tailed tests (one-tailed when predicted).

The sample includes 629 firm-year observations representing 38 COLI firms and 38 control firms from a sample matched on *ROA* and *lnEMP* in year $t-1$ (the year before adoption). Standard errors are calculated after clustering on firm.

Variable Definitions:

$FOREIGN_{t-1}$ = indicator variable equal to 1 if the absolute value of foreign pre-tax income (PIFO) divided by the absolute value of total pre-tax income (PI) is greater than 10 percent;

RD_{t-1} = indicator variable equal to 1 if R&D expense (XRD) is positive;

INDRANK = decile-ranking resulting from ranking all Compustat firms by year, industry, and total assets (AT);

SGRWTH = sales (SALE) in year t minus sales in year $t-1$, divided by sales in year $t-1$;

ETR_{t-1} = lagged value of total income taxes (TXT) divided by pre-tax book income (PI);

BODLINK2 = number of board interlocks the sample firm has in year t with firms that have adopted COLI prior to year t;

AUDITLINK2 = number of prior COLI adopters audited by the firm's auditor in year t, scaled by the total number of Compustat firms audited by the firm's auditor in year t;

INDLINK2 = number of prior COLI adopters in a firm's industry in year t, scaled by the total number of Compustat firms in the firm's industry in year t;

BEALINK2 = number of prior COLI adopters in the firm's BEA region in year t, scaled by the total number of Compustat firms in the firm's BEA region in year t; and

DISTANCE = ordinal variable equal to 1 for firms headquartered in the Mid-East BEA region, 2 for firms headquartered in the neighboring New England, Southeast, and Great Lakes BEA regions, 3 for firms headquartered in the Plains and Southwest BEA regions, and 4 for firms headquartered in the Rocky Mountain and Far West BEA region.

VI. SUMMARY AND CONCLUSION

This study contributes to the literature on aggressive corporate tax reporting by examining how the use of a particular corporate tax shelter, corporate-owned life insurance (COLI), spreads over time and across firms. I develop hypotheses based on four social environment factors derived from the literature on innovation-diffusion and mimetic isomorphism and test whether these factors help to explain the pattern of COLI adoptions. I find that COLI adopters are connected through board interlocks. Consistent with a cohesion model of diffusion, a direct tie to a prior COLI adopter via a board interlock increases the likelihood that a firm adopts the COLI shelter. I find no evidence that firms' connections to prior COLI adopters via a common auditor are related to the spread of COLI adoption. I also examine whether imitation of structurally equivalent peers helps to explain the spread of COLI participation. However, I find that the number of prior adopters in a firm's industry is not related to the likelihood that it will adopt COLI. Finally, I find some evidence that COLI use spreads geographically.

This study specifically examines which diffusing mechanisms help explain the spread of COLI use. Not all corporate practices diffuse the same way (Davis and Greve 1997), and the importance of a particular diffusion mechanism, like board ties, can wane over time (Mizruchi et al. 2006). Although the findings presented here do not necessarily generalize to all corporate tax shelters or to the adoption of tax strategies in today's information environment, the results related to COLI highlight the importance of examining firms' decisions in light of the broader social structures in which they operate. Moreover, the theories of innovation-diffusion and institutional isomorphism explored in this study are potentially applicable to a variety of accounting practice trends.

APPENDIX

Accounting Treatment of COLI

The level of detail, if any, disclosed in the financial statement footnotes regarding COLI plans varies considerably from firm to firm.[33] Investments in life insurance are governed by FASB Technical Bulletin No. 85-4, which prescribes the use of the cash surrender value (CSV) method. Under the CSV method, a firm should record the CSV of the life insurance policy as an asset, but under the right of setoff, firms can net outstanding policy loans against the CSV of the policy, rather than record a separate liability (APB No. 10, FASB Technical Bulletin No. 88-2, and FASB Interpretation No. 39). Given the right of setoff, firms can participate in the COLI shelter with little net effect on their balance sheets.[34]

As seen in the example below, other than through the tax effect, COLI has little income statement effect. In a leveraged COLI transaction, net COLI income (expense) equals increases in the cash surrender value (interest credited to the policy) plus any death benefits received less the premium expense and the interest expense on the policy loans (FASB Technical Bulletin No. 85-4). SFAS No. 109, *Accounting for Income Taxes*, indicates that the excess of CSV over premiums paid results in a permanent book-tax difference if the insurance policy is expected to be held until the death of the ensured. Thus, COLI programs that produce material book-tax differences should be reported in the firm's tax rate reconciliation. However, firms have considerable discretion over how permanent differences are netted and aggregated into specific line items on the rate reconciliation.

---

[33] Material for this section is drawn from Nurnberg (2004).
[34] Court documents reveal that Camelot's COLI plan was set to maintain a zero net equity balance at the end of each year (*In re C.M. Holdings, Inc.*, 254 B.R. 578 (D. Del. 2000).

	Book	Book-Tax Adjustment	Tax
Increase in policy cash surrender value	20	(20)	0
Less: Premium expense	(12)	12	0
Interest expense on policy loans	(10)	0	(10)
Net COLI income (expense)	(2)	(8)	(10)

Under IRC §72, the interest credited to the cash surrender value of the policy, $20 in the example above, is not included in the policyholder's gross income. Because proceeds from life insurance receive preferential tax treatment, policy premiums are not deductible for tax purposes. This example shows premium expense of $12, which reduces book income, but not taxable income. However, the interest expense paid on policy-backed loans, here $10, is tax deductible.

The numbers chosen for the example here are arbitrary, but show an important feature of the COLI shelter. The transaction produces a loss for tax purposes in excess of any economic or financial statement loss. Indeed, even though this example produces a $2 pre-tax book loss, at a 35 percent statutory rate, the $2.80 tax benefit from $8 of net nontaxable income offsets the $2 pre-tax book loss, generating positive after-tax income of $0.80.

REFERENCES

Abrahamson, E., and L. Rosenkopf. 1993. Institutional and competitive bandwagons: Using mathematical modeling as a tool to explore innovation diffusion. *Academy of Management Review* 18 (3): 487–517.

Allison, P. D. 1984. *Event History Analysis*. Beverly Hills, CA: Sage.

Bankman, J. 1999. The new market for corporate tax shelters. *Tax Notes* 83 (12): 1775–1795.

——— 2002. Commentary: Modeling the tax shelter world. *Tax Law Review* 55: 455–464.

——— 2004. An academic's view of the tax shelter battle. In *The Crisis in Tax Administration*, edited by H. J. Aaron, and J. Slemrod. Washington, D.C.: Brookings Institution.

Barth, M. E., W. H. Beaver, and W. R. Landsman. 1998. Relative valuation roles of equity book value and net income as a function of financial health. *Journal of Accounting and Economics* 25 (1): 1–34.

Berry, F. S., and W. D. Berry. 1990. State lottery adoptions as policy innovations: An event history analysis. *The American Political Science Review* 84: 395–415.

———, and ———. 1992. Tax innovation in the states: Capitalizing on political opportunity. *American Journal of Political Science* 36: 715–742.

———, and ———. 1994. The politics of tax increases in the states. *American Journal of Political Science* 38: 855–859.

Bizjak, J., M. Lemmon, and R. Whitby. 2009. Option backdating and board interlocks. *Review of Financial Studies* 22 (11): 4821–4847.

Box-Steffensmeier, J., and B. Jones. 1997. Time is of the essence: Event history models in political science. *American Journal of Political Science* 41 (4): 1414–1461.

Burns, L. R., and D. R. Wholey. 1993. Adoption and abandonment of matrix management programs: Effects of organizational characteristics and interorganizational networks. *Academy of Management Journal* 36 (1): 106–138.

Burt, R. S. 1987. Social contagion and innovation: Cohesion versus structural equivalence. *American Journal of Sociology* 92: 1287–1335.

Cyert, R., and J. G. March. 1963. *A Behavioral Theory of the Firm*. Englewood Cliffs, NJ: Prentice Hall.

Davis, G. F. 1991. Agents without principles? The spread of the poison pill through the intercorporate network. *Administrative Science Quarterly* 36: 605–633.

———, and H. R. Greve. 1997. Corporate elite networks and governance changes in the 1980s. *American Journal of Sociology* 103: 1–37.

Desai, M., and D. Dharmapala. 2006. Corporate tax avoidance and high-powered incentives. *Journal of Financial Economics* 79: 145–179.

DiMaggio, P. J., and W. W. Powell. 1983. The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review* 48: 147–160.

Eustice, J. 2002. Abusive tax shelters: Old "brine" in new bottles. *Tax Law Review* 55: 135–174.

Fligstein, N. 1985. The spread of the multidivisional form among large firms, 1919–1979. *American Sociological Review* 50: 377–391.

——— 1990. *The Transformation of Corporate Control*. Cambridge, MA: Harvard University Press.

Frank, M., L. Lynch, and S. Rego. 2009. Are financial and tax reporting aggressiveness reflective of broader corporate policies. *The Accounting Review* 84 (2): 467–498.

Galaskiewicz, J., and R. Burt. 1991. Interorganization contagion in corporate philanthropy. *Administrative Science Quarterly* 36 (1): 88–105.

Gergen, M. 2002. The logic of deterrence: corporate tax shelters. *Tax Law Review* 55: 255–286.

Graham, J. R., and A. L. Tucker. 2006. Tax shelters and corporate debt policy. *Journal of Financial Economics* 81: 563–594.

Gupta, S., and K. Newberry. 1997. Determinants of the variability in corporate effective tax rates: Evidence from longitudinal data. *Journal of Accounting and Public Policy* 16 (1): 1–34.

Han, S.-K. 1994. Mimetic isomorphism and its effect on the audit services market. *Social Forces* 73 (2): 637–664.

Haunschild, P. R. 1993. Interorganizational imitation: The impact of interlocks on corporate acquisition activity. *Administrative Science Quarterly* 38 (4): 564–592.

——— 1994. How much is that company worth? Interorganizational relationships, uncertainty and acquisition premiums. *Administrative Science Quarterly* 39: 391–412.

———, and A. S. Miner. 1997. Modes of interorganizational imitation: The effects of outcome salience and uncertainty. *Administrative Science Quarterly* 42 (3): 472–500.

Haveman, H. 1993. Follow the leader: Mimetic isomorphism and entry into new markets. *Administrative Science Quarterly* 38 (4): 593–627.

Lisowsky, P. 2010. Seeking shelter: Empirically modeling tax shelters using financial statement information. *The Accounting Review* 85 (5): 1693–1720.

Mahajan, V., and R. Peterson. 1985. *Models for Innovation Diffusion*. Beverly Hills, CA: Sage Publications.

March, J. G. 1981. Decisions in organizations and theories of choice. In *Perspectives on Organization Design and Behavior*, edited by Van de Ven, A., and W. F. Joyce. New York, NY: Wiley.

Mezias, S. 1990. An institutional model of organizational practice: Financial reporting at the Fortune 200. *Administrative Science Quarterly* 35 (3): 431–457.

Mintrom, M., and S. Vergari. 1998. Policy networks and innovation diffusion: The case of state education reforms. *The Journal of Politics* 60 (1): 126–148.

Mizruchi, M. S. 1992. *The Structure of Corporate Political Action*. Cambridge, MA: Harvard University Press.

———, L. B. Stearns, and C. Marquis. 2006. The conditional nature of embeddedness: A study of borrowing by large U.S. firms, 1973–1994. *American Sociological Review* 71 (2): 310–333.

Novack, J., and L. Saunders. 1998. The hustling of X-rated shelters. *Forbes* (December): 19.

Nurnberg, H. 2004. Accounting for company-owned life insurance. *Accounting Horizons* 18 (2): 109–126.

Omer, T. C., and M. K. Shelley. 2004. Competitive, political and economic factors influencing state tax policy changes. *The Journal of the American Taxation Association* 26 (Supp.): 103–126.

Palmer, D. A., P. D. Jennings, and X. Zhou. 1993. Late adoption of the multidivisional form by large U.S. corporations: Institutional, political, and economic accounts. *Administrative Science Quarterly* 38: 100–131.

Paull, L. 2000. JCT Testimony at Finance Committee hearing on Code's penalty and interest provisions. *Tax Notes Today* (March 9).

Pfeffer, J. 1981. Management as symbolic action: The creation and maintenance of organizational paradigms. *Research in Organizational Behavior* 3: 1–52.

Rao, H., and K. Sivakumar. 1999. Institutional sources of boundary-spanning structures: The establishment of investor relations departments in the Fortune 500 industrials. *Organization Science* 10 (1): 27–42.

Rogers, E. 2003. *Diffusion of Innovations*. 5th edition. New York, NY: Free Press.

Strang, D., and S. A. Soule. 1998. Diffusion in organizations and social movements: From hybrid corn to

poison pills. *Annual Review of Sociology* 24: 265–290.

Stuart, T. E., and S. Yim. 2010. Board interlocks and the propensity to be targeted in private equity transactions. *Journal of Financial Economics* 97: 174–189.

U.S. Department of the Treasury. 1990. Report to Congress on the taxation of life insurance company products. Washington, D.C.: U.S. Government Printing Office.

———. 1999. *The Problem of Corporate Tax Shelters: Discussion, Analysis, and Legislative Proposals*. Washington, D.C.: U.S. Government Printing Office.

U.S. Senate Permanent Subcommittee on Investigations. 2005. *The Role of Professional Firms in the U.S. Tax Shelter Industry*. S. Rep. 109-54. Washington, D.C.: U.S. Government Printing Office.

Weisbach, D. 2002. Ten truths about corporate tax shelters. *Tax Law Review* 55: 215–253.

Wilson, R. 2009. An examination of corporate tax shelter participants. *The Accounting Review* 84 (3): 969–999.

THE ACCOUNTING REVIEW
Vol. 86, No. 1
2011
pp. 59–100

American Accounting Association
DOI: 10.2308/accr.00000005

Voluntary Nonfinancial Disclosure and the Cost of Equity Capital: The Initiation of Corporate Social Responsibility Reporting

Dan S. Dhaliwal
The University of Arizona and Korea University

Oliver Zhen Li
The University of Arizona

Albert Tsang
Yong George Yang
The Chinese University of Hong Kong

ABSTRACT: We examine a potential benefit associated with the initiation of voluntary disclosure of corporate social responsibility (CSR) activities: a reduction in firms' cost of equity capital. We find that firms with a high cost of equity capital in the previous year tend to initiate disclosure of CSR activities in the current year and that initiating firms with superior social responsibility performance enjoy a subsequent reduction in the cost of equity capital. Further, initiating firms with superior social responsibility performance attract dedicated institutional investors and analyst coverage. Moreover, these analysts achieve lower absolute forecast errors and dispersion. Finally, we find that firms exploit the benefit of a lower cost of equity capital associated with the initiation of CSR disclosure. Initiating firms are more likely than non-initiating firms to raise equity capital following the initiations; among firms raising equity capital, initiating firms raise a significantly larger amount than do non-initiating firms.

Keywords: *corporate social responsibility; cost of capital; voluntary disclosure.*

Data Availability: *The data are publicly available from the sources identified in the paper.*

We appreciate the valuable comments from Ying Cao, Zhaoyang Gu, Michel Magnan, Morton Pincus, Suresh Radhakrishnan, T. J. Wong, and seminar participants at the City University of Hong Kong, Singapore Management University, Tsinghua University and 2009 AAA Annual Meeting, and especially the detailed constructive suggestions from Steven Kachelmeier (editor) and two anonymous referees. We thank Brian Bushee for generously sharing his classifications of institutional investors. We also thank CorporateRegister.com and KLD Research & Analytics, Inc. for providing the corporate social responsibility data, and I/B/E/S for the analyst following data. We are grateful to Xuemei Hao and Guangqing Li for their excellent research assistance. The work described in this paper was fully supported by a grant from the Research Grants Council of the Hong Kong SAR, China (Project No. CUHK451409).

Editor's note: Accepted by Steven Kachelmeier.

Submitted: April 2009
Accepted: May 2010
Published Online: January 2011

I. INTRODUCTION

The last 15 years have witnessed a steadily increasing emphasis on socially responsible corporate activities around the world. While third parties, such as KLD Research and Analytics, Inc. (KLD), often track and rate the corporate social responsibility (CSR) performance of large firms, firms have also become increasingly willing to voluntarily issue standalone CSR reports in recent years.[1] According to CorporateRegister.com, a private company that specializes in tracking CSR reports, few standalone CSR reports were issued in the United States before the mid-1990s. However, since then, increasingly more U.S. firms have committed to making this type of disclosure. In 2007 alone, large firms issued about 300 CSR reports. Although CSR disclosing firms represent only a small fraction of the population of U.S. publicly listed firms, their aggregate market value constituted over 10 percent of the total U.S. market capitalization in 2007.[2] The rapid increase in CSR reporting naturally raises questions among researchers: What are the rationales behind this type of voluntary disclosure? What benefits do firms gain by spending resources on compiling and publishing these standalone reports, especially given that CSR performance ratings are often available to investors through third parties?

A number of factors potentially provide answers to these questions, such as the growing influence of global enterprises, the intensified scrutiny of corporate impact on the society and the economy as a result of a loss of trust after a series of corporate scandals around 2001, and the recent rapid growth in ethical/socially responsible investment in the United States and around the world.[3] Anecdotal evidence also indicates that firms' reputations and long-term sales can suffer because of poor CSR performance. For instance, Nike struggled for years and invested a great amount of financial resources and effort to regain its reputation after the 1997 child labor scandal.[4]

We examine one factor, namely, a reduction in firms' cost of equity capital, that potentially provides an explanation for the increasing trend in CSR disclosure. Among various potential factors influencing CSR disclosure decisions, we focus on the cost of equity capital because it plays a critical role in a firm's financing and general operations decisions. Also, corporate executives appear to believe that voluntarily communicating information can reduce their firms' cost of capital (Graham et al. 2005). Further, there is a longstanding interest among academics in the relation between disclosure and the cost of capital (Diamond and Verrecchia 1991; Botosan 1997; Leuz and Verrecchia 2000; Botosan and Plumlee 2002).

To determine whether and how CSR disclosure is related to firms' cost of equity capital, we employ a sample of firms that intersect two CSR data sources: (1) a comprehensive list of firms releasing electronic or hard-copy standalone CSR reports since 1993, collected from various sources on the Internet; and (2) the KLD STATS database that provides detailed CSR performance ratings for individual firms. Our analyses provide four important insights. First, firms with a high cost of equity capital in the previous year are significantly more likely than others to initiate standalone CSR disclosures. Second, the cost of equity capital decreases for CSR-initiating firms with superior CSR performance. Third, CSR-initiating firms with superior CSR performance attract dedicated institutional investors and analyst coverage. Moreover, these analysts have more

---

1 Consistent with McWilliams and Siegel (2001), among others, we define CSR as instances where the company goes beyond compliance and voluntarily engages in actions that appear to advance social causes, including committing to environmental and human rights protection, providing community support, and so forth. In practice and academic research, CSR is often used interchangeably with "sustainability." We also follow this convention in the paper.

2 This figure is based on the mean market cap of $14.47 billion for firms as represented in Table 2 and the total U.S. market cap of around $15.35 trillion on May 23, 2007.

3 For example, according to the Social Investment Forum (2007), from 1995 to 2005, assets invested in socially responsible investment grew from $639 billion to $2.29 trillion, and accounted for approximately 11 percent of the total assets managed by professional managers.

4 See http://www.bandt.com.au/news/25/0c00d225.asp.

accurate forecasts and lower forecast dispersion. Finally, corroborating the result on the relation between CSR disclosure and the cost of equity capital, CSR-initiating firms are significantly more likely than non-initiating firms to conduct seasoned equity offerings (SEOs) in the two years following these initiations and among firms conducting SEOs, CSR-initiating firms raise a significantly larger amount of capital than do non-initiating firms. Overall, our evidence is consistent with our predictions that a potential reduction in the cost of equity capital motivates firms to publish standalone CSR reports and that CSR disclosure by firms with superior CSR performance leads to a lower cost of equity capital.

This study is the first to investigate the impact of standalone voluntary disclosure of general CSR issues on the cost of equity capital. We contribute to the literature by extending the traditional research on voluntary disclosure beyond the narrow focus of financial disclosure. The extant finance and accounting literatures on voluntary disclosure focus primarily on management forecasts or conference calls that are short-term-oriented.[5] In contrast, CSR disclosure, which is broad in scope, is related to a firm's long-term development strategies and performance sustainability. Our results provide evidence on the rationales behind and the consequences of the recent trend in voluntary CSR disclosure.

Our study is related to, but differs from, the work of Plumlee et al. (2008) and Richardson and Welker (2001). Plumlee et al. (2008) examine the impact of voluntary environmental disclosure quality on firm value. We examine a broader concept of CSR, which includes environmental protection, community development, corporate governance practices, employee relations, diversity practices, human rights, and product quality. In addition, we use a measure of CSR that is different from Plumlee et al. (2008), who use a self-constructed index to measure firms' environmental disclosure quality. We use a proxy that indicates whether firms publish CSR reports. Also, the information examined by Plumlee et al. (2008) comes from corporate environmental reports as well as annual reports and 10-Ks, which reflect both voluntary and mandatory disclosures. The standalone CSR reports we examine are voluntary.

Our study also differs from Richardson and Welker (2001), who examine the relation between the cost of equity capital and social as well as financial disclosure. First, we study U.S. firms, whereas they examine Canadian firms. The United States and Canada differ considerably in institutions related to information disclosure, with the United States having more stringent regulations than Canada (Richardson and Welker 2001). If more stringent regulations and the associated higher level of litigation risk translate into a generally higher level of disclosure credibility, then we can observe different relations between disclosure and the cost of equity capital in these two countries. In addition, the CSR measure used by Richardson and Welker (2001) is based on annual reports, whereas we focus on standalone CSR disclosures. These two forms of disclosure differ in depth and breadth of CSR coverage.

Methodologically, we differ from Plumlee et al. (2008) and Richardson and Welker (2001) by employing a lead-lag approach enhanced with two-stage regressions in sensitivity analysis to deal with endogeneity and self-selection issues and by exploring the underlying channels, such as institutional ownership and analyst coverage, through which CSR disclosure affects the cost of equity capital. In sum, we contribute to the literature by complementing and extending Plumlee et al. (2008) and Richardson and Welker (2001).

Section II develops our hypotheses. Section III describes our sample and methodology. Section IV presents empirical evidence on the relation between CSR disclosure and the cost of equity capital. Section V summarizes and concludes.

---

[5] One of the few exceptions is Dietrich et al. (2001), who investigate the effect of the supplemental disclosure of forward-looking information on security prices.

II. RELATED RESEARCH AND HYPOTHESIS DEVELOPMENT

Most prior research on the relation between disclosure and the cost of capital focuses on financial disclosure (Core 2001; Healy and Palepu 2001; Leuz and Wysocki 2008). The consensus appears to be that a negative relation exists between the quality of financial disclosure and the cost of capital. Greater disclosure increases investors' awareness of a firm's existence and enlarges its investor base, which improves risk-sharing and reduces the cost of capital (Merton 1987). In addition, higher quality or more precise firm-specific disclosures decrease the covariance of a firm's cash flow with the cash flows of other firms (Hughes et al. 2007; Lambert et al. 2007), which essentially reduces the betas of individual firms and, hence, the cost of equity capital. Similarly, greater disclosure can lead to reduced information asymmetry among investors or between managers and investors. When the level of disclosure is inadequate and some investors are perceived to be better informed than others, informationally disadvantaged investors price-protect themselves and become less willing to trade. The resultant illiquidity increases the bid-ask spread and transaction costs (Verrecchia 2001), which leads to a higher required rate of return or cost of equity capital (Amihud and Mendelson 1986).

These mechanisms likely apply to both financial and nonfinancial disclosure, as long as the information concerned is value-relevant. Indeed, a fair amount of research suggests that CSR information is value-relevant (Margolis and Walsh 2001; Orlitzky et al. 2003; Al-Tuwaijri et al. 2004). Of course, CSR practices can affect firms' financial performance and value through channels other than those related to financial disclosure. For instance, voluntary socially responsible behavior can help firms avoid government regulation and, therefore, reduce compliance costs. In addition, socially responsible firms appeal to consumers who care about the corresponding social issues, which leads to superior sales and financial performance (Lev et al. 2010). Socially aware investors are willing to pay a premium for the securities of socially responsible firms (Anderson and Frankel 1980; Richardson and Welker 2001). Perhaps more important, some CSR projects have direct implications for positive cash flow even in the near future. For example, practices related to protecting the environment and improving employee welfare can reduce potential litigation and pollution cleaning costs, boost employee morale. And, thereby, production efficiency. These arguments highlight the importance of CSR disclosure in reducing information asymmetry and uncertainty related to factors affecting firm value (Rodriguez et al. 2006), which in turn reduces the cost of equity capital.

Nevertheless, a straightforward generalization of the cost of capital effect from financial disclosure to nonfinancial CSR disclosure is not always obvious. Standalone CSR reports are currently subject to very limited regulatory guidance. There is a common concern about the usefulness of this type of disclosure because of noncomparability and potential credibility issues and opportunistic behaviors of firms (Ingram and Frazier 1980; Hobson and Kachelmeier 2005).[6] In the end, whether *voluntary* CSR disclosure reduces a firm's cost of equity capital is an empirical question.

It is important to note that CSR performance ratings of large firms are often available to investors through third parties. These ratings could be directly associated with the cost of equity capital of these firms. However, ratings alone are unlikely to provide sufficient information for investors to assess firms' overall CSR performance. Detailed CSR disclosures potentially provide additional information necessary for investors to assimilate these summary ratings.[7] Further, voluntarily disclosing CSR activities demonstrates firms' confidence in their CSR performance, which

---

6 Although some accounting and consulting firms provide voluntary assurance service (Simnett et al. 2009), there is not yet a government standard that regulates this service, and the assurance industry is still in its infancy.

7 An obvious analogy is the usefulness of footnote disclosures and management discussions in supplementing financial statements.

sends a positive signal to investors, or, in the case of poor CSR performance, allows firms to offer explanations. Therefore, CSR disclosures contain information beyond that contained in CSR performance ratings.

Some firms also disclose information on CSR activities in their annual reports or filings with the SEC. However, a firm's voluntary compilation and publication of standalone CSR reports demonstrates its special effort and commitment to improving transparency regarding long-term performance and risk management. More importantly, compared with the CSR information provided in annual reports or 10-Ks, standalone CSR reports are more comprehensive and contain significantly more details.[8] Therefore, standalone CSR reports likely provide incrementally useful information for investors to evaluate firms' long-term sustainability. Focusing on standalone CSR reports can thus improve the power of our tests and shed light on this new form of voluntary nonfinancial disclosure.

Our first hypothesis predicts that a possible reduction in the cost of equity capital provides an incentive for firms to publish CSR reports. Frankel et al. (1995) find that firms increase their level of voluntary disclosure to raise capital in the future at a lower cost, which suggests that firms with a relatively higher cost of capital likely have a greater incentive to enhance disclosure. Lending support to the cost of capital incentive for disclosure, Sletten (2008) finds that stock price declines, which imply an increase in firms' cost of equity capital, induce managers to disclose more information.[9]

Of course, endogeneity and self-selection issues can arise if we examine a contemporaneous relation between CSR disclosure and the cost of equity capital. On the one hand, if CSR disclosure is motivated by a firm's desire to reduce its high cost of equity capital, then we should find a positive relation between CSR disclosure and the cost of the equity capital. On the other hand, if CSR disclosure leads to a lower cost of equity capital, then we should find a negative relation between CSR disclosure and the cost of equity capital. Therefore, the contemporaneous relation between CSR disclosure and the cost of equity capital could be ambiguous. To address the potential endogeneity and self-selection issues related to CSR disclosure and the cost of equity capital, we employ a lead-lag approach in our main analyses and state our first hypothesis below:

H1: The likelihood that a firm will disclose its corporate social responsibility activities is positively associated with its cost of equity capital in the previous year.

If CSR disclosure provides information that is incremental to information provided in third-party CSR performance ratings or other information dissemination channels such as annual reports or 10-Ks, then the preceding discussion suggests that CSR disclosure should lead to a lower cost of equity capital. This logic suggests the following hypothesis:

H2: Corporate social responsibility disclosure is associated with a subsequently lower cost of equity capital.

Support for H1 and H2 would provide justification for the rationales behind and the consequences of CSR disclosure. We also test a corollary of H1 and H2 by examining whether disclos-

---

[8] In untabulated analyses and relying on manual data collection, we compare CSR-related content in the first-time standalone CSR reports and annual reports (or 10-Ks in the absence of annual reports) of 50 firms out of our final sample of 213 firms. We find that, on average, standalone CSR reports are significantly longer (28.3 pages versus 1.5 pages) and cover significantly more CSR issues (6.4 issues versus 1.5 issues) compared to annual reports or 10-Ks. The inference of the above comparison is also supported by a comprehensive survey conducted by KPMG (2008), which finds that among the largest 100 U.S. firms, only about 1 percent of them adequately integrate CSR reports into their annual reports.

[9] However, the result documented by Sletten (2008) could be attributable to either the numerator (cash flow) effect or the denominator (cost of capital) effect, or both.

ing firms seek external financing after CSR disclosures. If CSR disclosure is motivated by firms' desire to reduce the cost of equity capital, then these firms will be more likely than non-disclosing firms to raise equity capital after their CSR disclosures to exploit the reduction in their cost of equity capital, and they will also strive to raise a larger amount. While we formulate our predictions based on CSR disclosure, in the empirical analysis we focus on CSR-disclosure-initiating firms since initial reports likely contain more information than mundane continuing reports.

III. SAMPLE AND METHODOLOGY

Sample Description

CSR disclosure policies can be sticky across years. Therefore, we focus on first-time standalone CSR reports. We collect standalone CSR reports issued by U.S. firms from various sources, including (1) Corporate Social Responsibility Newswire, (2) CorporateRegister.com, (3) Internet searches, and (4) company websites. The first two sources are the two leading organizations collecting and disseminating news and information related to CSR. We verify our CSR reporting sample by checking whether we can find their actual standalone CSR reports.[10]

In our main analyses, we control for the relative social responsibility performance of sample firms, as proxied for by the KLD social performance rating scores. Our final sample comprises firms that are in both the KLD STATS and Compustat databases. KLD evaluates CSR performance for all covered firms along a variety of dimensions, regardless of whether they release standalone reports.[11] Starting from 1991, KLD STATS rated approximately 650 companies every year, comprising mainly all firms in the S&P 500 and Domini 400 Social SM Index. During 2001 to 2002, KLD expanded its coverage to include the largest 1,000 U.S. companies by market capitalization. Since 2003, it has covered the largest 3,000 U.S. companies based on market capitalization.

Table 1, Panel A shows the industry distribution, based on Barth et al.'s (1998) industry classifications, of CSR reports and disclosing firms. During the 1993–2007 period, 294 firms issued a total of 1,190 standalone CSR reports.[12] The Utilities industry has the largest proportion (30.4 percent) of firms publishing CSR reports, while the Services and Insurance/Real Estate industries have the lowest proportion of disclosing firms (2.15 percent and 0.40 percent, respectively). Consistent with the broad scope of CSR disclosure, many non-pollution-prone industries including the Food and Retail industries also actively disclose their social performance. After eliminating 81 firms because of missing data, our final sample contains 213 disclosing firms. The Utilities industry constitutes the largest proportion of the final sample (13.4 percent). Table 1, Panel B presents the distribution by year of CSR reports and disclosing firms. Overall, there is a steadily increasing trend in the number of CSR reports over time from 8 in 1993 to 184 in 2007. The average report length nearly doubles from about 20 pages in the early 1990s to more than 40 pages in the most recent years. On average, a CSR report has 36 pages.[13]

---

[10] It is tempting to examine the information content of CSR disclosures. However, this test is hampered by the lack of information on the exact reporting dates of the reports. Nevertheless, we conduct an event study based on the reporting months of the reports. We find that (1) during the CSR reporting month, there is no difference in raw and market-adjusted returns between high and low CSR performance firms; (2) during the three-month period following the CSR reporting month, high CSR performance firms appear to do slightly better than low CSR performance firms, based on market-adjusted returns; and (3) there is no difference in returns between CSR reporting months and non-CSR reporting months.

[11] The Appendix to this paper lays out the main categories of CSR issues employed by KLD in its rating process and also the average rating scores across industries.

[12] Sometimes a firm publishes multiple CSR reports, often discussing different CSR-related issues such as environmental versus non-environmental matters, in a single year. When that is the case, we combine them into one firm-year observation.

[13] The statistics for page numbers are based on all CSR reports published in the year, not just on first-time reports.

TABLE 1

Sample Distribution

Panel A: Distribution by Industry

	Industries	No. of CSR Reports	%	No. of CSR Reporters	%	No. of Firms in KLD Database	% of KLD Firms Disclosing CSR Reports	Average No. of Pages per Report	No. of First-Time CSR Reports	%
		Initial Sample							Full Sample	
1	Mining/Construction	33	2.77	9	3.06	83	10.84	43	7	3.29
2	Food	73	6.13	17	5.78	86	19.77	43	14	6.57
3	Textiles/Print/Publish	75	6.30	20	6.80	168	11.90	32	16	7.51
4	Chemicals	102	8.57	18	6.12	94	19.15	25	12	5.63
5	Pharmaceuticals	80	6.72	14	4.76	297	4.71	38	12	5.63
6	Extractive	96	8.07	16	5.44	180	8.89	33	13	6.10
7	Manf: Rubber/glass/etc.	18	1.51	4	1.36	53	7.55	47	3	1.41
8	Manf: Metal	28	2.35	4	1.36	100	4.00	30	3	1.41
9	Manf: Machinery	30	2.52	9	3.06	120	7.50	34	8	3.76
10	Manf: Electrical Eqpt	22	1.85	6	2.04	133	4.51	21	4	1.88
11	Manf: Transport Eqpt	62	5.21	13	4.42	87	14.94	41	10	4.69
12	Manf: Instruments	32	2.69	9	3.06	232	3.88	35	8	3.76
13	Manf: Misc.	3	0.25	2	0.68	25	8.00	53	2	0.94
14	Computers	123	10.34	27	9.18	593	4.55	36	23	10.80
15	Transportation	30	2.52	12	4.08	275	4.36	41	8	3.76
16	Utilities	188	15.80	45	15.31	148	30.41	39	29	13.62
17	Retail: Wholesale	22	1.85	5	1.70	106	4.72	24	3	1.41
18	Retail: Misc.	29	2.44	12	4.08	222	5.41	38	10	4.69
19	Retail: Restaurant	17	1.43	4	1.36	48	8.33	52	3	1.41
20	Financial	92	7.73	34	11.56	778	4.37	33	18	8.45
21	Insurance/Real Estate	1	0.08	1	0.34	248	0.40	29	0	0.00
22	Services	14	1.18	8	2.72	372	2.15	22	5	2.35
23	Others	20	1.68	5	1.70	25	20.00	48	2	0.94
	Total	1,190	99.99	294	99.97				213	100.01

(continued on next page)

Panel B: Distribution by Year

Year	No. of CSR Reports	%	No. of First-Time CSR Reporters	%	No. of Firms in KLD Database	% of KLD Firms Disclosing CSR Reports	Average No. of Pages per Report	No. of First-Time CSR Reports	%
	Initial Sample							Final Sample	
1993	8	0.67	6	2.04	432	1.85	22	3	1.41
1994	16	1.34	11	3.74	423	3.78	24	7	3.29
1995	23	1.93	10	3.40	440	5.23	25	5	2.35
1996	31	2.61	12	4.08	456	6.80	23	9	4.23
1997	46	3.87	14	4.76	459	10.02	23	11	5.16
1998	41	3.45	8	2.72	494	8.30	23	4	1.88
1999	56	4.71	15	5.10	518	10.81	25	7	3.29
2000	57	4.79	18	6.12	537	10.61	23	11	5.16
2001	101	8.49	42	14.29	928	10.88	28	34	15.96
2002	99	8.32	28	9.52	981	10.09	31	18	8.45
2003	101	8.49	22	7.48	2,639	3.83	33	14	6.57
2004	121	10.17	20	6.80	2,750	4.40	40	16	7.51
2005	149	12.52	32	10.88	2,792	5.34	42	30	14.08
2006	157	13.19	31	10.54	2,790	5.63	41	23	10.80
2007	184	15.46	25	8.50	2,827	6.51	45	21	9.86
Total	1,190	100.01	294	99.97				213	100.00

This table provides the sample distribution by industry and year for both the initial CSR reports collected and CSR reports selected for the final sample. "No. of First-Time CSR Reports" is the number of CSR reports in the earliest reporting year of a firm (i.e., CSR report-initiating year) based on our collected CSR reports. "% of KLD Firms Disclosing CSR Reports" is the number of CSR disclosing firms in any industry (namely, "No. of CSR Reporters" in Panel A) or year (namely, "No. of CSR Reports" in Panel B) divided by the total number of KLD firms in that industry or year, respectively.

Empirical Models and Variable Definitions

Past Cost of Equity Capital and Current-Year CSR Disclosure

To test H1, we examine whether a high cost of equity capital in the previous year gives firms an incentive for CSR disclosure in the current year. In the empirical regression model, we control for other determinants of CSR disclosure to parse out potential confounding effects. However, the current literature provides limited information on what motivates a firm's CSR disclosure decision. As CSR disclosure is part of a firm's overall voluntary disclosure strategy, we identify potential factors from the voluntary disclosure literature that influence a firm's decision to commit to CSR disclosure. Our logistic regression model is specified as follows:

$$\begin{aligned}\log[\text{prob}(DISCI_{i,t})/(1-\text{prob}(DISCI_{i,t}))] = {} & \beta_0 + \beta_1 COC_{i,t-1} + \beta_2 PERFORM_{i,t-1} \\ & + \beta_3 HICONCERN_{i,t-1} + \beta_4 SIZE_{i,t-1} \\ & + \beta_5 LITIGATION_{i,t-1} + \beta_6 ROA_{i,t-1} \\ & + \beta_7 COMPETITION_{i,t-1} + \beta_8 FIN_{i,t-1} \\ & + \beta_9 TOBINQ_{i,t-1} + \beta_{10} LEV_{i,t-1} + \beta_{11} GLOBAL_{i,t-1} \\ & + \beta_{12} LIQUIDITY_{i,t-1} + \beta_{13} ABS\_EM_{i,t-1} \\ & + \beta_{14} CIG_{i,t-1} + \Sigma IND_{i,t} + \Sigma YEAR_{i,t} + \varepsilon_{i,t} \qquad (1)\end{aligned}$$

where $DISCI_{i,t}$ is an indicator variable that equals 1 if firm *i* discloses a standalone CSR report for the first time in year *t* (initiating firm-years or initiators), and 0 (non-initiating firm-years or non-initiators) otherwise. Therefore, the control group (*DISCI* = 0), namely, non-initiators, includes all years of firms that never issue CSR reports and the years before and after CSR-initiating firms' first-time reports.

Our main variable of interest, the cost of equity capital in the year prior to first-time CSR disclosure, *COC*, is the *ex ante* or implied cost of equity capital, calculated using three different models, namely, those of Gebhardt et al. (2001), Claus and Thomas (2001), and Easton (2004). The mean of the three measures (*COC_AVG*) serves as our proxy for the cost of equity capital. To implement the estimation, we obtain expected future earnings per share from I/B/E/S and market price and dividend per share from Compustat.

We include a number of control variables in the regression. *PERFORM* is the total KLD score of CSR strengths, which we use to proxy for firms' CSR performance. Firms with better social performance have a greater incentive to disclose (Dye 1985). The KLD database is widely used in CSR research (Graves and Waddock 1994; Berman et al. 1999; Baron et al. 2009). Waddock (2003, 369) regards it as "the *de facto* (CSR) research standard at the moment."[14] KLD ranks firms' CSR performance in seven main categories: (1) community, (2) corporate governance, (3) diversity, (4) employee relations, (5) environment, (6) human rights, and (7) product.[15] We adjust raw CSR strength scores each year by industry medians to get relative performance scores that are comparable across industries.

---

[14] Of course, there is also no lack of criticism of the KLD database. For example, KLD uses indicator variables to describe firms' CSR performance. This is a crude methodology and potentially suffers from loss of information. Chatterji et al. (2009) show that KLD environmental strengths do not accurately predict pollution levels or compliance violations, and that KLD ratings do not optimally use publicly available data.

[15] The rankings are based on information obtained from surveys, financial statements, government documents, peer-reviewed legal journals, and reports from mainstream media. KLD defines a set of potential strengths under each category and assigns a value of 1 if a strength exists, and a value of 0 otherwise. See the Appendix for more details on KLD's rating categories.

We control for firm size (*SIZE*) because size captures various factors motivating firms to issue CSR reports such as public pressure or financial resources (Lang and Lundholm 1993). We measure *SIZE* as the natural logarithm of the market value of common equity at the beginning of each year. Skinner (1997) argues that firms facing a higher level of litigation risk (*LITIGATION*) are more likely to make voluntary disclosure to preempt potential lawsuits. *LITIGATION* is an indicator variable that equals 1 if a firm operates in a high-litigation industry (SIC codes of 2833–2836, 3570–3577, 3600–3674, 5200–5961, and 7370), and 0 otherwise (Francis et al. 1994; Matsumoto 2002). As firms with better financial performance likely have more resources to practice CSR activities and produce CSR reports, we include return on assets (*ROA*), computed as income before extraordinary items scaled by total assets at the beginning of each year.

Dye (1985) suggests that proprietary costs arising from product market competition can reduce disclosure incentives. Hence, we control for industry competition (*COMPETITION*), which is proxied by the Herfindahl-Hirschman Index multiplied by -1. This index is computed as the sum of the squared fractions of sales of the 50 largest firms in an industry (industries are defined based on the two-digit SIC codes). In cases where there are fewer than 50 firms in an industry, we use all firms in the industry to calculate market shares. In addition, firms raising capital in the public market have a greater propensity to make voluntary disclosures (Frankel et al. 1995). We control for a firm's financing activities (*FIN*) by assessing the amount of debt or equity capital raised by the firm during the year scaled by total assets at the beginning of the year. Following Richardson et al. (2004), *FIN* is measured as the sale of common and preferred shares minus the purchase of common and preferred shares plus the long-term debt issuance minus the long-term debt reduction.

We also control for growth opportunities (*TOBINQ*) because firms in an expansionary period are more financially constrained and have fewer resources for CSR activities and disclosure. However, growth firms also tend to have higher levels of information asymmetry, which could induce managers to make more disclosures to attract potential investors. The net effect is hence unknown *ex ante*. *TOBINQ* is Tobin's Q, defined as the market value of common equity plus the book value of preferred stock, book value of long-term debt and current liabilities, scaled by the book value of total assets. We include the debt ratio (*LEV*) in the model because debt servicing plays a monitoring role and debt holders demand greater disclosure (Leftwich et al. 1981). We define *LEV* as the ratio of total debt divided by total assets.

In addition, firms with a global focus, especially those operating in emerging markets, face greater pressure to commit to social performance and are accordingly more likely to provide CSR disclosure. *GLOBAL* is an indicator variable that equals 1 if a firm reports foreign income, and 0 otherwise. Further, managers have incentives to increase the liquidity of their firms' stock in order to issue equities or sell shares of their firm obtained from options or other incentive compensation plans. One way to increase liquidity is to improve transparency and supply more information to investors. Our liquidity measure, *LIQUIDITY*, is the ratio of the number of shares traded in the year to the total shares outstanding at the year-end.

Finally, CSR disclosure could be correlated with the general disclosure policies and financial transparency of firms. To control for this possibility, we include two variables to proxy for firm financial disclosure quality and voluntary disclosure policy: earnings quality (*ABS_EM*) and management earnings forecasts (*CIG*). We use the absolute value of abnormal accruals from the modified Jones (1991) model, based on Dechow et al. (1995), to proxy for earnings quality (Francis et al. 2008).[16] Following prior studies that use management forecasts as a direct measure

[16] Using the original Jones (1991) model or an alternative version developed by Dechow et al. (2003, 359, Equation (2b)) yields similar results.

of a firm's disclosure policy (Rogers and Van Buskirk 2009), we define *CIG* as an indicator variable that equals 1 if a firm issues at least one earnings forecast in the year, and 0 otherwise. In all specifications of the model, we include industry and year indicators to control for potential industry and year effects.

Effect of CSR Disclosure on the Future Cost of Equity Capital

Hypothesis 2 predicts that CSR disclosure leads to a lower cost of equity capital. We test H2 by estimating the following regression model:

$$\Delta\%COC_{i,t+1} = \beta_0 + \beta_1 DISCI_{i,t} + \beta_2\Delta SIZE_{i,t} + \beta_3\Delta BETA_{i,t} + \beta_4\Delta LEV_{i,t} + \beta_5\Delta MB_{i,t} + \beta_6\Delta LTG_{i,t} + \beta_7\Delta LNDISP_{i,t} + \Sigma IND_{i,t} + \Sigma YEAR_{i,t} + \varepsilon_{i,t} \quad (2)$$

where $\Delta\%COC_{i,t+1}$ is the percentage change in the cost of equity capital from year t to year t+1. The control variables also adopt the change form. A negative coefficient on *DISCI* would support H2.

The control variables are derived from prior research. Fama and French (1992) find that expected returns are negatively associated with firm size and positively associated with the book-to-market ratio. Hence, we include firm size (*SIZE*) and the market-to-book ratio (*MB*). The market model *BETA*, which is estimated using CRSP daily data for each year, is included to control for systematic risk. Gebhardt et al. (2001) and Gode and Mohanram (2003) find that the implied cost of equity capital is positively associated with long-term growth rate. We therefore include an empirical proxy of long-term growth rate based on I/B/E/S analyst EPS forecasts (*LTG*), which is measured as the difference between the two-year-ahead consensus EPS forecast and the one-year-ahead consensus EPS forecast scaled by the one-year-ahead consensus EPS forecast. Gebhardt et al. (2001) and Dhaliwal et al. (2005) find that analyst forecast dispersion is negatively associated with the implied cost of equity capital. Thus, we include analyst forecast dispersion (*LNDISP*), which is calculated as the logarithm of the standard deviation of analyst EPS forecasts divided by the consensus forecast. We include leverage (*LEV*) because Fama and French (1992) suggest that the cost of equity capital increases as the degree of leverage increases. All other variables are as defined earlier.

Although firms may be motivated by a possible reduction in the cost of equity capital when deciding whether to issue a CSR report, from the perspective of investors, CSR disclosure *per se* may not necessarily warrant a lower cost of equity capital. Corporate managers could attempt to manage public impressions through such disclosures; therefore, CSR information can be self-serving and noncredible (Cormier and Magnan 2003; Hobson and Kachelmeier 2005). Investors are likely to have a favorable perception if a firm actually performs well in its CSR practices relative to its peers. To incorporate this possibility, we augment Equation (2) with a measure of a firm's relative CSR performance from KLD (*HIPERFORM*):

$$\Delta\%COC_{i,t+1} = \beta_0 + \beta_1 DISCI_{i,t} + \beta_2 HIPERFORM_{i,t} + \beta_3 DISCI_{i,t} * HIPERFORM_{i,t} + \beta_4\Delta SIZE_{i,t} + \beta_5\Delta BETA_{i,t} + \beta_6\Delta LEV_{i,t} + \beta_7\Delta MB_{i,t} + \beta_8\Delta LTG_{i,t} + \beta_9\Delta LNDISP_{i,t} + \varepsilon_{i,t} \quad (3)$$

where *HIPERFORM* is an indicator variable that equals 1 if a firm's CSR performance score, *PERFORM*, is higher than its industry median (in other words, if the firm is a superior CSR performer in its industry), and 0 otherwise. All other variables are as defined earlier. We expect the effect of *DISCI* ∗ *HIPERFORM* to be negative. In an additional test, instead of using the interaction term between *DISCI* and *HIPERFORM*, we estimate Equation (2) within the high and low partitions of CSR performance scores. Using partitioned subsamples sacrifices some power due to

reduced sample size, but has the benefit of flexibility that allows the effects of other variables to also vary based on high or low levels of CSR performance scores.

Endogeneity and self-selection could potentially affect our results. In our main analysis, we use a lead-lag approach to tackle these issues. To further enhance inferences based on our lead-lag approach, we adopt the Heckman and Hausman two-stage procedures and repeat our main analyses. The Heckman two-stage procedure introduces the inverse Mills ratio into the second-stage OLS regression to control for self-selection bias that is related to CSR disclosure. We obtain qualitatively similar results using the Heckman two-stage procedure.[17] The Hausman test deals with potential endogeneity in the data. We conduct the Hausman test and find that endogeneity does not qualitatively affect our main results.

IV. RESULTS

Descriptive Statistics

Table 2, Panel A provides descriptive statistics for the variables included in Equation (1) for the full sample and separately for initiators and non-initiators. The cost of equity capital before CSR disclosure is significantly higher ($p = 0.04$) among CSR initiators (12.86 percent) than among non-initiators (11.98 percent). This difference is also reflected in a significantly positive correlation coefficient between *DISCI* and *COC_AVG* in Table 2, Panel B, providing initial support for H1.

Consistent with the theory on voluntary disclosure, firms voluntarily publishing standalone CSR reports tend to have superior CSR performance (*PERFORM*) relative to their industry peers. The difference in CSR performance between the two groups (1.613 for initiators versus −0.166 for non-initiators) is significant ($p < 0.01$). The correlation between *DISCI* and *PERFORM* is also significantly positive though at a relatively moderate level of 0.09 based on the Spearman correlation and 0.13 based on the Pearson correlation (see Table 2, Panel B). This highlights the importance of including *PERFORM* in our regression equations.

Initiators are significantly larger (*SIZE*: 9.147 for initiators versus 5.783 for non-initiators, $p < 0.01$) and more profitable (*ROA*: 0.051 for initiators versus 0.015 for non-initiators, $p < 0.01$) than non-initiators, lending support to the financial resources argument for CSR disclosure. Contrary to the proprietary information argument, initiators tend to observe greater industry competition than non-initiators (*COMPETITION*: −0.060 for initiators versus −0.069 for non-initiators, $p = 0.02$).

Initiators have a significantly lower level of financing than non-initiators (*FIN*: −0.019 for initiators versus 0.043 for non-initiators, $p < 0.01$). The negative financing level for initiators implies that these firms, in net effect, either have repurchased stock or redeemed their debts. Firms normally conduct repurchases when they believe that their stock is undervalued, indicating a high cost of equity capital, which in turn provides an incentive for managers to increase disclosure and transparency levels. Similarly, the redemption of mature debts likely implies that firms need future financing to maintain a normal capital level. These firms would also be willing to increase their level of disclosure if doing so helped them to lower the cost of borrowing.

Initiators have a higher degree of leverage than non-initiators (*LEV*: 0.265 for initiators versus 0.221 for non-initiators, $p < 0.01$). Those with a higher level of global operations are also more likely to publish CSR reports (*GLOBAL*: 0.460 for initiators versus 0.219 for non-initiators, $p < 0.01$), consistent with the notion that these firms attract more attention in the international community. Contrary to the notion of disclosing information to improve liquidity, initiators actually

[17] The only exception is for the test of analyst forecast errors. Among better CSR performers, the coefficient on *DISCI* is positive and insignificant.

TABLE 2

Summary Statistics and Correlation

Panel A: Mean Comparison

Variable	Full Sample	$DISCI_t = 1$ (n = 213)	$DISCI_t = 0$ (n = 11,712)	t-value (difference)
COC_AVG_{t-1}(%)	11.988	12.855	11.985	2.08
$PERFORM_{t-1}$	−0.142	1.613	−0.166	11.66
$SIZE_{t-1}$	5.795	9.147	5.783	41.13
$LITIGATION_{t-1}$	0.204	0.199	0.204	−0.22
ROA_{t-1}	0.017	0.051	0.015	13.06
$COMPETITION_{t-1}$	−0.069	−0.060	−0.069	2.28
FIN_{t-1}	0.042	−0.019	0.043	−14.36
$TOBINQ_{t-1}$	2.046	1.994	2.047	−0.53
LEV_{t-1}	0.221	0.265	0.221	4.70
$GLOBAL_{t-1}$	0.192	0.460	0.219	5.44
$LIQUIDITY_{t-1}$	1.248	1.387	1.247	2.10
CIG_{t-1}	0.528	0.646	0.527	15.09
ABS_EM_{t-1}	0.068	0.032	0.066	−6.84

Panel B: Spearman\Pearson correlation (n (*DISCI* = 0) = 11,712, n (*DISCI* = 1) = 213)

	$DISCI_t$	COC_AVG_{t-1}	$PERFORM_{t-1}$	$SIZE_{t-1}$	$LITIGATION_{t-1}$	ROA_{t-1}	$COMPETITION_{t-1}$
$DISCI_t$		**0.01**	**0.13**	**0.11**	0.00	**0.02**	**0.01**
COC_AVG_{t-1}	**0.01**		0.00	**0.01**	**−0.07**	**−0.11**	**0.03**
$PERFORM_{t-1}$	**0.09**	0.01		**0.23**	0.00	**0.07**	**−0.03**
$SIZE_{t-1}$	**0.08**	**0.01**	**0.21**		**−0.02**	**0.25**	**0.09**
$LITIGATION_{t-1}$	0.00	**−0.09**	**−0.03**	**−0.03**		**−0.16**	**0.10**
ROA_{t-1}	**0.02**	**−0.06**	**0.04**	**0.24**	**−0.03**		**−0.09**
$COMPETITION_{t-1}$	**0.01**	**0.07**	**−0.03**	**0.12**	**0.07**	**−0.18**	
FIN_{t-1}	**−0.02**	**−0.11**	**−0.06**	**−0.14**	**0.08**	**−0.09**	**0.02**

(continued on next page)

Panel B: Spearman\Pearson correlation (n (*DISCI* = 0) = 11,712, n (*DISCI* = 1) = 213)

	$DISCI_t$	COC_AVG_{t-1}	$PERFORM_{t-1}$	$SIZE_{t-1}$	$LITIGATION_{t-1}$	ROA_{t-1}	$COMPETITION_{t-1}$
$TOBINQ_{t-1}$	0.00	**−0.32**	**0.02**	**0.23**	**0.18**	**0.29**	**0.10**
LEV_{t-1}	**0.02**	**0.15**	0.01	**0.08**	**−0.15**	**−0.15**	**−0.03**
$GLOBAL_{t-1}$	**0.05**	**−0.05**	**0.03**	**0.24**	**0.03**	**0.10**	**−0.07**
$LIQUIDITY_{t-1}$	**0.02**	**−0.06**	−0.01	**0.28**	**0.22**	**0.05**	**−0.04**
CIG_{t-1}	**0.08**	**0.02**	**0.08**	**0.23**	**0.04**	**0.10**	0.00
ABS_EM_{t-1}	**−0.03**	**−0.07**	**−0.02**	**−0.21**	**0.05**	**−0.04**	**−0.06**

	FIN_{t-1}	$TOBINQ_{t-1}$	LEV_{t-1}	$GLOBAL_{t-1}$	$LIQUIDITY_{t-1}$	CIG_{t-1}	ABS_EM_{t-1}
$DISCI_t$	**−0.02**	0.00	**0.01**	**0.03**	**0.01**	**0.08**	**−0.03**
COC_AVG_{t-1}	**−0.09**	**−0.22**	**0.14**	0.01	**−0.01**	**0.01**	**−0.04**
$PERFORM_{t-1}$	**−0.06**	**0.02**	0.01	**0.03**	**−0.02**	**0.09**	**−0.04**
$SIZE_{t-1}$	**−0.14**	**0.15**	**0.04**	**0.12**	**0.22**	**0.22**	**−0.20**
$LITIGATION_{t-1}$	**0.11**	**0.17**	**−0.13**	**0.01**	**0.21**	**0.04**	**0.05**
ROA_{t-1}	**−0.28**	**−0.15**	**−0.06**	**0.04**	**−0.05**	**0.08**	**−0.20**
$COMPETITION_{t-1}$	**0.03**	**0.08**	**−0.04**	**0.02**	**0.04**	**0.02**	**−0.02**
FIN_{t-1}		**0.30**	**−0.27**	**−0.04**	**0.08**	**−0.07**	**0.14**
$TOBINQ_{t-1}$	**0.18**		**−0.25**	**−0.03**	**0.29**	**0.03**	**0.21**
LEV_{t-1}	**−0.26**	**−0.37**		**−0.01**	**−0.06**	0.00	**−0.05**
$GLOBAL_{t-1}$	**−0.05**	**0.01**	**−0.01**		**0.04**	**0.07**	**−0.05**
$LIQUIDITY_{t-1}$	**0.08**	**0.32**	**−0.07**	**0.12**		**0.17**	**0.14**
CIG_{t-1}	**−0.06**	**0.08**	**0.01**	**0.14**	**0.22**		**−0.06**
ABS_EM_{t-1}	**0.10**	**0.14**	**−0.11**	**−0.07**	**0.11**	**−0.06**	

All continuous variables are winsorized at the 1st and 99th percentiles. In Panel B, the Spearman (Pearson) correlations are below (above) the diagonal. A correlation coefficient in bold indicates that the correlation is statistically significant at least at the 10 percent level.

Variable Definitions:

$DISCI_t$ = indicator variable that equals 1 if it is the earliest reporting year of a firm that issues CSR reports (CSR report initiating year), and 0 otherwise;

COC_AVG_{t-1} = implied cost of equity capital (in percentage) in year $t-1$ estimated as the mean of three different models: Gebhardt et al. (2001), Claus and Thomas (2001), and Easton (2004);

(continued on next page)

$PERFORM_{t-1}$ = measure of social performance defined as the industry-adjusted total CSR strength scores from the seven CSR rating categories obtained from the KLD Research & Analytics database;

$SIZE_{t-1}$ = natural logarithm of the market value of equity (CSHO * PRCC_F) at the beginning of each year;

$LITIGATION_{t-1}$ = indicator variable that equals 1 if the firm operates in a high-litigation industry (SIC codes of 2833–2836, 3570–3577, 3600–3674, 5200–5961, and 7370), and 0 otherwise;

ROA_{t-1} = total return on assets measured as the ratio of income before extraordinary items (IB) over total assets (AT) at the beginning of each year;

$COMPETITION_{t-1}$ = Herfindahl-Hirschman Index multiplied by −1. The Herfindahl-Hirschman Index is calculated by summing the squares of the market shares of the 50 largest companies in an industry. We calculate a firm's market share by dividing the sales (SALE) of a firm in year t by the total sales of all of the 50 largest companies in an industry in that year. We define industries based on the two-digit SIC codes. In cases where there are fewer than 50 companies in an industry, we use all companies in that industry to calculate the market share of each firm;

FIN_{t-1} = amount of debt or equity capital raised by the firm scaled by total assets at the beginning of year t–1. It is measured as the issuance of common stock and preferred shares minus the purchase of common stock and preferred shares (SSTK – PRSTKC) plus the long-term debt issuance minus the long-term debt reduction (DLTIS – DLTR);

$TOBINQ_{t-1}$ = market value of common equity plus the book value of preferred stock (PSTKL), book value of long-term debt (DLTT) and current liability (LCT), scaled by the book value of total assets;

LEV_{t-1} = leverage ratio, which is defined as the ratio of total debt (DLTT + DLC) divided by total assets;

$GLOBAL_{t-1}$ = indicator variable that equals 1 if the firm reports non-zero foreign income (PIFO), and 0 otherwise;

$LIQUIDITY_{t-1}$ = ratio of the number of shares traded in year t–1 to the total shares outstanding at the end of year t–1;

CIG_{t-1} = indicator variable that equals 1 if the firm issues an earnings forecast in year t–1, and 0 otherwise. We obtain data for *CIG* from First Call database; and

ABC_EM_{t-1} = absolute value of abnormal accruals estimated based on the modified Jones model. The modified-Jones model discretionary accrual is an estimated cross-sectional each year using all firm-year observations in the same two-digit SIC code as follows: $TA_{i,t} = \beta_0 + \beta_1(1 / Assets_{i,t-1}) + \beta_2(\Delta Sales_{i,t} - \Delta REC_{i,t}) + \beta_3 PPE_{i,t} + \varepsilon_{i,t}$, where TA_{it} is defined as the change in non-cash current assets minus the change in current liabilities excluding the current portion of long-term debt, minus depreciation and amortization, scaled by lagged total assets; $\Delta Sales_{it}$ is the change in sales scaled by lagged total assets; $\Delta REC_{i,t}$ is the change in accounts receivable scaled by lagged total assets; and PPE_{it} is net property, plant, and equipment scaled by lagged total assets. Only industry-years with at least eight firms for estimation are considered. We use the residuals from the annual cross-sectional industry regression model above as discretionary accruals.

have higher liquidity levels than non-initiators (*LIQUIDITY*: 1.387 for initiators versus 1.247 for non-initiators, $p = 0.04$). Finally, initiators have better financial disclosure as manifested in their more frequent management forecasts (*CIG*: 0.646 for initiators versus 0.527 for non-initiators, $p < 0.01$) and better earnings quality (*ABS_EM*: 0.032 for initiators versus 0.066 for non-initiators, $p < 0.01$) than non-initiators.

Cost of Equity Capital and the Likelihood of CSR Disclosure

Hypothesis 1 predicts that a firm's likelihood of disclosing its corporate social responsibility activities is positively associated with its cost of equity capital in the previous year. We report the regression results for Equation (1) in Table 3. In Column I, we include all first-time reporting firm-year observations. In Column II, we exclude first-time reports that primarily discuss environmental issues, following Simnett et al. (2009). In Column III, we examine the robustness of our results to the exclusion of the Utilities industry.

Across all three specifications of the dependent variable, the cost of equity capital, *COC_AVG*, in year $t-1$, is significantly positively associated with a firm's likelihood of voluntarily issuing a standalone CSR report in year t (coeff. = 0.049, $p < 0.01$; coeff. = 0.052, $p < 0.01$; and coeff. = 0.062, $p < 0.01$ in Columns I, II and III, respectively), consistent with H1, which posits that a higher past cost of equity capital is associated with a greater likelihood of voluntary CSR disclosure in the current year. In Column I, for instance, holding other factors constant, when the prior year cost of equity capital increases by one percentage point, the odds of initiating standalone CSR disclosure increase by 5.02 percent.

The coefficient estimates of the control variables are generally consistent with the univariate comparisons in Table 2. One exception is *LIQUIDITY*, which reverses direction. The significantly negative coefficient suggests that firms with lower levels of liquidity are more likely to publish CSR reports, consistent with our original conjecture. The effects of financial disclosure quality, *ABS_EM*, and management forecast, *CIG*, are no longer significant.

CSR Disclosure and the Future Cost of Equity Capital

Hypothesis 2 predicts that voluntary CSR disclosure leads to a lower future cost of equity capital. Table 4, Panel A compares initiators and non-initiators and Table 4, Panel B presents the regression results. In Column I (Equation (2)), the coefficient on *DISCI* is insignificant (coeff. = -0.037, $p > 0.50$). It appears that CSR disclosure *per se* is not significantly associated with a change in a firm's future cost of equity capital. In Column II (Equation (3)), we consider whether a firm has superior CSR performance relative to its industry peers. The interaction term between *DISCI* and *HIPERFORM* is significantly negative (coeff. = -4.618, $p < 0.01$), consistent with H2, which posits that CSR disclosure reduces the cost of equity capital.[18] Combining the main effect of *DISCI* and the effect of the interaction term between *DISCI* and *HIPERFORM* in Column II, we infer that superior CSR performers enjoy a 1.833 percent reduction in the cost of equity capital when they produce standalone CSR reports for the first time. In Columns III and IV, we obtain similar results when we exclude environmental reports and the Utilities industry, respec-

---

[18] We perform a sensitivity test by restricting the analysis to firm-year observations of CSR reporters in a pre-post setting. Specifically, we focus only on disclosing firms and still use the change specification of the dependent variable Δ*COC_AVG*. *DISCI* is an indicator variable that equals 1 for the first reporting year and equals 0 before or after the first reporting year of a disclosing firm. The purpose of this examination is to show that a reduction in the cost of equity capital occurs immediately after the first reporting year and to alleviate the concern that the size of initiator sample is small relative to the universe of firm-year observations used in the main test. We obtain similar results and inferences, namely, firms with superior CSR performance enjoy a reduction in the cost of equity capital if they publish standalone CSR reports.

TABLE 3

Determinants of CSR Disclosure

Dependent Variable = $DISCI_t$		I		II		III	
		Full Sample		Removing Environmental Reports		Removing Utilities	
Variables	Pred. Sign	Coeff.	Prob.	Coeff.	Prob.	Coeff.	Prob.
COC_AVG_{t-1}	+	0.049***	0.00	0.052***	0.00	0.062***	0.00
$PERFORM_{t-1}$	+	0.101**	0.03	0.057	0.25	0.100**	0.04
$SIZE_{t-1}$	+	0.967***	0.00	1.017***	0.00	1.001***	0.00
$LITIGATION_{t-1}$	?	0.208	0.45	0.154	0.59	0.187	0.50
ROA_{t-1}	+	1.404	0.47	1.407	0.49	1.805	0.36
$COMPETITION_{t-1}$	−	1.685	0.37	1.105	0.55	1.569	0.40
FIN_{t-1}	+	0.768	0.50	0.955	0.41	1.406	0.20
$TOBINQ_{t-1}$	?	−0.398***	0.00	−0.355***	0.01	−0.371***	0.00
LEV_{t-1}	+	−0.846	0.28	−0.851	0.30	−1.038	0.20
$GLOBAL_{t-1}$	+	0.078*	0.08	0.091**	0.05	0.084*	0.07
$LIQUIDITY_{t-1}$	−	−0.292**	0.02	−0.255**	0.04	−0.292**	0.02
CIG_{t-1}	+	0.313	0.16	0.324	0.18	0.272	0.27
ABS_EM_{t-1}	−	−2.217	0.49	−1.545	0.64	−3.137	0.36
Year Indicators		Yes		Yes		Yes	
Industry Indicators		Yes		Yes		Yes	
Pseudo R^2		0.258		0.244		0.271	
Likelihood Ratio		303.40		27.00		273.30	
n (dep. var. = 1)		213		164		182	
n		11,925		11,876		11,419	

*, **, *** Indicate that the estimated coefficient is statistically significant at the 10 percent, 5 percent, and 1 percent levels, respectively.

This table presents the logistic regression results with $DISCI_t$ in year t as the dependent variable, while all control variables are in year t−1. All continuous variables are winsorized at the 1st and the 99th percentiles. All t-statistics are corrected using the Huber-White procedure.

All variables are defined as in Table 2.

TABLE 4
Post-CSR Disclosure Cost of Equity Capital

Panel A: Mean Comparison

	Full Sample	$DISCI_t = 1$	$DISCI_t = 0$	t-value (difference)
ΔCOC_AVG_{t+1} (%)	1.691	1.393	1.693	−0.90
$\Delta SIZE_t$	−0.003	−0.002	−0.003	0.13
$\Delta BETA_t$	−0.019	0.061	−0.020	2.79
ΔLEV_t	−0.023	−0.027	−0.023	−4.01
ΔMB_t	−0.015	−0.012	−0.015	0.85
ΔLTG_t	0.004	−0.051	0.004	−1.10
$\Delta LNDISP_t$	0.023	0.009	0.023	−1.47

Panel B: Post-CSR Disclosure Cost of Equity Capital (Dependent Variable = ΔCOC_AVG_{t+1})

	I		II		III		IV	
			Full Sample		Removing Environmental Reports		Removing Utilities	
Variables	Coeff.	t-stat.	Coeff.	t-stat.	Coeff.	t-stat.	Coeff.	t-stat.
$DISCI_t$	−0.037	−0.06	2.785***	2.54	4.699***	3.56	1.856	1.45
$HIPERFORM_t$			0.030	0.20	0.041	0.27	0.052	0.33
$DISCI_t * HIPERFORM_t$			−4.618***	−3.38	−6.015***	−3.74	−4.045***	−2.61
$\Delta SIZE_t$	4.698***	3.70	4.849***	3.82	4.814***	3.79	5.040***	3.86
$\Delta BETA_t$	0.342***	3.53	0.331***	3.40	0.334***	3.44	0.298***	2.99
ΔLEV_t	−7.220***	2.86	−7.107***	2.81	−5.116**	−2.04	−6.557***	−2.54
ΔMB_t	−10.137***	−14.19	−10.190***	−14.23	−10.257***	−14.33	−10.107***	−13.86
ΔLTG_t	−0.175**	−2.45	−0.173**	−2.42	−0.170**	−2.37	−0.175**	−2.37
$\Delta LNDISP_t$	0.111	0.30	0.107	0.29	0.109	0.30	0.243	0.64
Year Indicators	Yes		Yes		Yes		Yes	
Industry Indicators	Yes		Yes		Yes		Yes	
Adjusted R^2	0.062		0.064		0.062		0.062	

(continued on next page)

Panel B: Post-CSR Disclosure Cost of Equity Capital (Dependent Variable = ΔCOC_AVG_{t+1})

	I		II		III		IV	
			Full Sample		Removing Environmental Reports		Removing Utilities	
Variables	**Coeff.**	**t-stat.**	**Coeff.**	**t-stat.**	**Coeff.**	**t-stat.**	**Coeff.**	**t-stat.**
n ($DISCI_t = 1$)	213		213		164		184	
n ($DISCI_t = 1$ and $HIPERFORM_t = 1$)			165		127		150	
n	9,254		9,254		9,205		8,820	

Panel C: Post-CSR Disclosure Cost of Equity Capital Partitioned Conditional on Firm's CSR Performance

	Top 50% *PERFORM.*		Bottom 50% *PERFORM*	
Variables	**Coeff.**	**t-stat.**	**Coeff.**	**t-stat.**
$DISCI_t$	−1.777**	−2.04	2.130	1.41
$\Delta SIZE_t$	3.200***	6.40	5.673***	3.99
$\Delta BETA_t$	0.360***	3.12	0.239	0.79
ΔLEV_t	−4.857*	−1.68	−7.747	−1.11
ΔMB_t	−12.258***	−14.16	−4.778**	−2.13
ΔLTG_t	−0.332***	−3.87	−0.478**	−2.17
$\Delta LNDISP_t$	0.414	0.93	−0.960	−0.78
Year Indicators	Yes		Yes	
Industry Indicators	Yes		Yes	
Adjusted R^2	0.048		0.059	
n ($DISCI = 1$)	153		60	
n	4,691		4,563	

*, **, *** Indicate that the estimated coefficient is statistically significant at the 10 percent, 5 percent and 1 percent levels, respectively.

All continuous variables are winsorized at the 1st and 99th percentiles. All t-statistics are corrected using the Huber-White procedure. Analyst forecast data are obtained from I/B/E/S.

Variable Definitions:

$DISCI_t$ = indicator variable that equals 1 if it is the earliest reporting year of a firm that issues CSR reports (CSR report initiating year), and 0 otherwise;

ΔCOC_AVG_{t+1} = change in cost of equity capital (in percentage) from year t to year t+1;

(continued on next page)

$HIPERFORM_t$ = indicator variable that equals 1 if the firm's total CSR strength score, namely, *PERFORM* in year *t*, is higher than the industry median (in other words, if the firm is classified as a high CSR performer in its industry), and 0 otherwise;

$\Delta SIZE_t$ = change in size where *SIZE* is measured as the natural logarithm of the market value of equity (CSHO * PRCC_F) at the beginning of each year;

$\Delta BETA_t$ = change of beta from year t–1 to year t where *BETA* is estimated from the market model using the daily CRSP stock returns;

ΔLEV_t = change in leverage from year t–1 to year t where *LEV* is the leverage ratio, which is defined as the ratio of total debt (DLTT + DLC) divided by total assets;

ΔMB_t = change of market-to-book ratio from year t–1 to year t where the market-to-book ratio is defined as the market value of equity divided by book value of equity (CEQ);

ΔLTG_t = long-term growth rate where *LTG* is measured as the difference between the mean two-year-ahead analyst consensus EPS forecast and the mean one-year-ahead analyst consensus EPS forecast divided by the mean one-year-ahead analyst consensus EPS forecast; and

$\Delta LNDISP_t$ = change in analyst forecast dispersion from year t–1 to year t, where *LNDISP* is measured as the logarithm of the standard deviation of analyst estimates of year t earnings divided by the consensus forecast of year t earnings.

tively. Overall, the evidence is consistent with our H2 that CSR-disclosing firms with superior CSR performance achieve a reduction in the cost of equity capital.[19]

Table 4, Panel C presents the results from estimating Equation (2) within the two subsamples partitioned based on annual industry medians of CSR performance (*PERFORM*). Consistent with the results in Panel B, we find a significantly negative coefficient on *DISCI* (coeff. = −1.777, p = 0.04) in the high CSR performance subsample. This coefficient indicates that voluntary CSR disclosure yields a 1.77 percent reduction in the cost of equity capital. In the low CSR performance subsample, there is no significant association between CSR disclosure and the change in the cost of equity capital.[20]

Potential Mechanisms Linking CSR Disclosure and the Cost of Equity Capital

The above results suggest that CSR disclosure combined with superior CSR performance is associated with a reduction in the cost of equity capital. Below, we provide evidence on the potential underlying mechanisms through which voluntary CSR disclosure lowers the cost of equity capital. We focus on two types of financial intermediaries: institutional investors and financial analysts.

CSR Disclosure and Institutional Investors

Shleifer and Vishny (1986) suggest that the large equity stakes in the invested firms and the high levels of sophistication of these investors enable them to reduce agency cost problems and the extent of information asymmetry between managers and shareholders, an effect that leads to a reduction in the cost of equity capital. We consider three different types of institutional investors: dedicated (*DED*), transient (*TRA*), and quasi-indexer (*QIX*) institutional investors. Dedicated institutional investors are more likely to play monitoring and governance roles than the other two types (Bushee 1998). To determine whether CSR disclosure attracts institutional investors, we follow Bushee and Noe (2000) and estimate the following model:

$$\begin{aligned}\Delta INST_{i,t+1} = {} & \beta_0 + \beta_1 DISCI_{i,t} + \beta_2 HIPERFORM_{i,t} + \beta_3 DISCI_{i,t} * HIPERFORM_{i,t} + \beta_4 INST_{i,t-1} \\ & + \beta_5 \Delta MRET_{i,t} + \beta_6 TVOL_{i,t-1} + \beta_7 \Delta MV_{i,t} + \beta_8 BETA_{i,t-1} + \beta_9 IRISK_{i,t-1} \\ & + \beta_{10} \Delta LEV_{i,t} + \beta_{11} \Delta DP_{i,t} + \beta_{12} \Delta EP_{i,t} + \beta_{13} \Delta MB_{i,t} + \beta_{14} \Delta SGR_{i,t} + \beta_{15} \Delta RATE_{i,t} \\ & + \beta_{16} \Delta SHRS_{i,t} + \Sigma IND_{i,t} + \Sigma YEAR_{i,t} + \varepsilon_{i,t}, \qquad (4)\end{aligned}$$

[19] In alternative specifications of the model, we examine the effects of two other variables proxying for firms' effort and commitment to better CSR disclosure. (1) We identify firms that provide assurance (*ASSURANCE*) of their reports through independent third parties, most often Big 4 accounting firms and international consulting companies. Simnett et al. (2009) provide evidence that firms seeking to enhance the creditability of their reports and their corporate reputation are more likely to have their sustainability reports assured. (2) We also assess the effect of the length of each CSR report (*LENGTH*) relative to the average report length of the disclosing firm's industry (Leuz and Schrand 2008). Of course, *ASSURANCE* and *LENGTH* are not independent of *DISCI*. We find that, conditional on first-time CSR disclosure (*DISCI*), external assurance and long report length further reduce the cost of equity capital. Specifically, when we use the *ASSURANCE* indicator (equals 1 with an assurance, and 0 otherwise), the coefficient on *DISCI* ∗ *HIPERFORM* is negative and significant (coeff. = −3.523, $p < 0.01$) and the coefficient on *DISCI* ∗ *HIPERFORM* ∗ *ASSURANCE* is negative and significant (coeff. = −3.540, $p = 0.08$). Therefore, assurance doubles the effect of CSR disclosure. When we use the *LENGTH* indicator (equals 1 if longer than the industry-year median, and 0 otherwise), the coefficient on *DISCI* ∗ *HIPERFORM* is negative and significant (coeff. = −2.574, p = 0.02) and the coefficient on *DISCI* ∗ *HIPERFORM* ∗ *LENGTH* is negative and significant (coeff. = −3.930, p = 0.06). Therefore, a long report more than doubles the effect of CSR disclosure.

[20] The coefficient on *DISCI* is not significant but positive, if anything, for poor CSR performers. It is possible that disclosing poor CSR performance could actually be a signal of high risk or firm weakness and, therefore, the cost of equity capital could actually go up. This does explain why the direct effect of *DISCI* is insignificant in the pooled regression.

where Δ denotes a change from year t to year $t+1$. *INST* represents stock ownership by dedicated (*DED*), transient (*TRA*), or quasi-indexer (*QIX*) institutional investors. *MRET* is the market-adjusted buy-and-hold stock return measured over the year. *TVOL*, a liquidity proxy, is the average monthly trading volume relative to total shares outstanding. *IRISK* is the logarithmic transformation of the standard deviation of market-model residuals calculated using daily stock returns. Beta (*BETA*), debt ratio (*LEV*), and *IRISK* capture firm risk along different dimensions. *DP* is the ratio of dividends to the market value of equity. *EP* is the ratio of income before extraordinary items to the market value of equity. *SGR* is the percentage change in annual sales. We include *DP*, *EP*, *MB*, and *SGR* to control for changes in firms' fundamentals that can affect the investment decisions of institutional investors (Bushee 2001). *RATE* is the S&P stock rating (9 = A+, 8 = A, 7 = A–, 6 = B+, 5 = B, 4 = B−, 3 = C, 2 = D, 1 = not rated), which captures the preference of institutional investors for well-reputed firms (Del Guercio 1996). *SHRS* is the logarithmic transformation of shares outstanding, and its change form proxies for equity issuance or repurchases that affect both institutional investor following and firms' disclosure policies. All other variables are as defined earlier.

Table 5, Panel A presents comparisons of one-year-ahead holdings and changes in holdings by the three types of institutional investors between initiators and non-initiators. Overall, the univariate comparisons do not reveal significant differences between initiators and non-initiators. If anything, we observe a greater decrease in transient institutional holding among initiators compared to non-initiators ($p = 0.04$), even though the level of this type of holding is still slightly higher among initiators than among non-initiators ($p = 0.07$).

Table 5, Panel B displays the regression results. There is weak evidence that initiating firms with superior CSR performance attract more dedicated institutional investors. The coefficient on *DISCI* $*$ *HIPERFORM* is marginally significantly positive (coeff. = 0.414, $p = 0.16$). To further examine this issue, we run regressions without the interaction term in the two subsamples partitioned based on annual industry medians of CSR performance for dedicated institutional investors. We report the results in Table 5, Panel C. We observe a significantly positive coefficient on *DISCI* (coeff. = 0.438, $p = 0.01$) for the superior-performance group, whereas the coefficient on *DISCI* for the low-performance group is insignificant. In untabulated tests, we do not find a significant association between transient or quasi-indexer institutional investor holdings and the initiation of CSR disclosure for the full sample or the partitioned subsamples.

In sum, the evidence in this subsection suggests that voluntary CSR disclosure attracts dedicated institutional investors, who have long investment horizons and play monitoring and governance roles. Consistent with our previous evidence that superior CSR performers enjoy a reduction in the cost of equity capital through CSR disclosure, the effect of CSR disclosure on dedicated institutional ownership is stronger if disclosing firms have CSR performance superior to their industry peers.

CSR Disclosure and Analyst Forecasts

We also examine three questions related to financial analysts and CSR disclosure. First, we explore whether financial analysts are more willing to cover firms after they initiate CSR disclosure. Second, we investigate whether the level of forecast accuracy increases and finally we determine whether forecast dispersion decreases when CSR reports are available. Increased levels of analyst coverage and forecast accuracy and a reduction in the level of forecast dispersion have the potential to lower the cost of equity capital. To determine the impact of CSR disclosure on the behavior of financial analysts, we run the following three regressions following Lang and Lundholm (1996) and Ali et al. (2007):

TABLE 5

Post-CSR Disclosure Changes in Institutional Ownership

Panel A: Mean Comparison

	Full Sample	$DISCI_t = 1$	$DISCI_t = 0$	t-value (difference)
$DED_{t+1}(\%)$	11.996	11.285	12.019	−0.79
$TRA_{t+1}(\%)$	16.507	18.572	16.440	1.82
$QIX_{t+1}(\%)$	29.809	30.014	29.802	0.24
ΔDED_{t+1}	−0.551	−0.995	−0.536	−0.57
ΔTRA_{t+1}	−0.452	−2.293	−0.390	−2.07
ΔQIX_{t+1}	1.018	1.947	0.986	1.00

Panel B: Post-CSR Disclosure Institutional Investor Holdings

	ΔDED_{t+1}		ΔQIX_{t+1}		ΔTRA_{t+1}	
Variables	**Coeff.**	**t-stat.**	**Coeff.**	**t-stat.**	**Coeff.**	**t-stat.**
$DISCI_t$	0.190	0.80	0.267	0.63	0.053	0.16
$HIPERFORM_t$	−0.037	−0.76	−0.129	−1.44	0.085	1.23
$DISCI_t * HIPERFORM_t$	0.414	1.40	−0.093	−0.17	−0.052	−0.13
DED_{t-1}	−0.018***	−5.94				
QIX_{t-1}			−0.032***	−6.75		
TRA_{t-1}					−0.059***	−15.37
$\Delta MRET_t$	0.043	0.82	−0.107	−1.07	−0.010	−0.13
$TVOL_{t-1}$	0.019	0.57	0.033	0.49	0.066	1.24
ΔMV_t	0.071	0.74	0.026	0.14	−0.149	−1.08
$BETA_{t-1}$	−0.053	−0.78	−0.010	−0.08	−0.082	−0.85
$IRISK_{t-1}$	0.056	0.69	0.065	0.42	−0.087	−0.74
ΔLEV_t	−0.008	−0.02	−0.924	−1.40	0.324	0.63
ΔDP_t	4.032	1.52	11.935**	2.39	6.617**	1.69
ΔEP_t	0.169	0.61	0.152	0.29	−0.356	−0.88

(continued on next page)

Panel B: Post-CSR Disclosure Institutional Investor Holdings

Variables	ΔDED_{t+1} Coeff.	ΔDED_{t+1} t-stat.	ΔQIX_{t+1} Coeff.	ΔQIX_{t+1} t-stat.	ΔTRA_{t+1} Coeff.	ΔTRA_{t+1} t-stat.
ΔMB_t	−0.010	−1.14	0.011	0.66	0.026**	1.96
ΔSGR_t	−0.001	−0.22	0.002	0.87	−0.001	−0.18
$\Delta RATE_t$	0.034***	2.79	0.037*	1.66	0.023	1.35
$\Delta SHRS_t$	0.298**	2.34	−0.109	−0.47	−0.287	−1.60
Year Indicators	Yes		Yes		Yes	
Industry Indicators	Yes		Yes		Yes	
Adjusted R^2	0.747		0.792		0.724	
n ($DISCI_t = 1$)	170		170		170	
n ($DISCI_t = 1$ and $HIPERFORM_t = 1$)	126		126		126	
n	9,342		9,342		9,342	

Panel C: Post-CSR Disclosure Institutional Investor Holdings Conditional on Firm's CSR Performance

	TOP 50% *PERFORM*		Bottom 50% *PERFORM*	
	ΔDED_{t+1}		ΔDED_{t+1}	
Variables	**Coeff.**	**t-stat.**	**Coeff.**	**t-stat.**
$DISCI_t$	0.438***	2.67	−0.063	−0.24
DED_{t-1}	−0.017***	−3.91	−0.027***	−6.45
$\Delta MRET_t$	0.047	0.60	−0.031	−0.42
$TVOL_{t-1}$	0.028	0.53	0.066	1.41
ΔMV_t	0.145	1.01	0.050	0.38
$BETA_{t-1}$	−0.151	−1.57	0.039	0.39
$IRISK_{t-1}$	0.178	1.57	−0.126	−1.06
ΔLEV_t	0.472	0.90	−0.199	−0.42
ΔDP_t	5.020	1.27	3.690	1.01
ΔEP_t	−0.104	−0.26	0.427	1.08
ΔMB_t	−0.012	−0.97	0.009	0.68

(continued on next page)

Panel C: Post-CSR Disclosure Institutional Investor Holdings Conditional on Firm's CSR Performance

Variables	TOP 50% *PERFORM* ΔDED_{t+1} Coeff.	t-stat.	Bottom 50% *PERFORM* ΔDED_{t+1} Coeff.	t-stat.
ΔSGR_t	−0.004**	−2.42	0.001	1.00
$\Delta RATE_t$	0.042***	2.49	0.008	0.43
$\Delta SHRS_t$	0.457***	2.55	0.130	0.71
Year Indicators	Yes		Yes	
Industry Indicators	Yes		Yes	
Adjusted R^2	0.519		0.504	
n ($DISCI_t = 1$)	126		44	
n	4,685		4,657	

*, **, *** Indicate that the estimated coefficient is statistically significant at the 10 percent, 5 percent, and 1 percent levels, respectively.

All continuous variables are winsorized at the 1st and 99th percentiles. All t-statistics are corrected using the Huber-White procedure.

Panel A presents the OLS regression results with ΔDED_{t+1}, ΔQIX_{t+1}, or ΔTRA_{t+1} as the dependent variable.

Variable Definitions:

$DISCI_t$ = indicator variable that equals 1 if it is the earliest reporting year of a firm that issues CSR reports (CSR report initiating year), and 0 otherwise;

DED_{t+1}, QIX_{t+1}, and TRA_{t+1} = percentage of ownership holding by dedicated, quasi-indexer, and transient institutional investors, respectively, relative to total shares outstanding in year $t+1$;

ΔDED_{t+1}, ΔQIX_{t+1}, and ΔTRA_{t+1} = changes in percentage of ownership from year t to year $t+1$. If there is a missing value on the percentage holding of any type of institutional investor when a firm's total institutional investor holding does not equal 0, then we assume a value of 0;

$HIPERFORM_t$ = indicator variable that equals 1 if the firm's total CSR strength score, namely, *PERFORM*, is higher than the industry median (in other words, if the firm is classified as a high CSR performer in its industry), and 0 otherwise;

$\Delta MRET_t$ = $MRET_t - MRET_{t-1}$; where *MRET* is the market-adjusted buy-and-hold stock return measured over the year of interest with a minimum of 125 observations;

$TVOL_{t+1}$ = average monthly trading volume relative to total shares outstanding measured over the year of interest;

ΔMV_t = $MV_t - MV_{t-1}$; where *MV* is the logarithm of the market value of equity (CSHO * PRCC_F);

$BETA_{t-1}$ = market-model beta calculated from the daily stock returns in year $t-1$ with a minimum of 125 observations;

$IRISK_{t-1}$ = logarithm of the standard deviation of the market-model residuals calculated from the daily stock returns in year $t-1$ with a minimum of 125 observations;

ΔLEV_t = change in leverage from year $t-1$ to year t where *LEV* is the leverage ratio, which is defined as the ratio of total debt (DLTT + DLC) divided by total assets;

ΔDP_t = $DP_t - DP_{t-1}$; where *DP* is the ratio of dividends (DVC) to the market value of equity;

(continued on next page)

$\Delta EP_t = EP_t - EP_{t-1}$; where *EP* is the ratio of income before extraordinary items (IB) to the market value of equity;
$\Delta MB_t = MB_t - MB_{t-1}$; where *MB* is the market-to-book value of equity, which is defined as the market value of equity divided by the book value of equity (CEQ);
$\Delta SGR_t = SGR_t - SGR_{t-1}$; where *SGR* is the percentage change in annual sales (SALE);
$\Delta RATE_t = RATE_t - RATE_{t-1}$; where *RATE* is the S&P stock rating (9 = A+, 8 = A, 7 = A−, 6 = B+, 5 = B, 4 = B−, 3 = C, 2 = D, 1 = not rated); and
$\Delta SHRS_t = SHRS_t - SHRS_{t-1}$; where *SHRS* is the logarithm of shares outstanding.

$$\Delta COVERAGE_{i,t+1} = \beta_0 + \beta_1 DISCI_{i,t} + \beta_2 HIPERFORM_{i,t} + \beta_3 DISCI_{i,t} * HIPERFORM_{i,t} + \beta_4 \Delta SIZE_{i,t} + \beta_5 \Delta STDROE_{i,t} + \beta_6 \Delta INVPRICE_{i,t} + \beta_7 \Delta RETVAR_{i,t} + \beta_8 \Delta RD_{i,t} + \beta_9 \Delta ROA_{i,t} + \beta_{10} \Delta CORR_{i,t} + \varepsilon_{i,t}, \quad (5)$$

$$\Delta|FE|_{i,t+1} = \beta_0 + \beta_1 DISCI_{i,t} + \beta_2 HIPERFORM_{i,t} + \beta_3 DISCI_{i,t} * HIPERFORM_{i,t} + \beta_4 \Delta SIZE_{i,t} + \beta_5 \Delta STDROE_{i,t} + \beta_6 ACHEPS_{i,t} + \beta_7 \Delta RD_{i,t} + \beta_8 \Delta ROA_{i,t} + \beta_9 \Delta CORR_{i,t} + \varepsilon_{i,t}, \quad (6)$$

$$\Delta DISP_{i,t+1} = \beta_0 + \beta_1 DISCI_{i,t} + \beta_2 HIPERFORM_{i,t} + \beta_3 DISCI_{i,t} * HIPERFORM_{i,t} + \beta_4 \Delta SIZE_{i,t} + \beta_5 \Delta STDROE_{i,t} + \beta_6 ACHEPS_{i,t} + \beta_7 \Delta RD_{i,t} + \beta_8 \Delta ROA_{i,t} + \beta_9 \Delta CORR_{i,t} + \varepsilon_{i,t}, \quad (7)$$

where *COVERAGE* is the 12-month average of the number of analysts who issue annual earnings forecasts captured in the I/B/E/S database for a specific firm; $|FE|$ is the absolute value of the 12-month average of analyst forecast errors, which is defined as actual earnings minus the mean forecast, deflated by the stock price at the beginning of the fiscal year; and *DISP* is the 12-month average of the standard deviations of analyst forecasts, deflated by the stock price at the beginning of the fiscal year.

We include a number of control variables derived from prior research. We include firm size (*SIZE*) because larger firms have more potential brokerage or investment banking businesses for analysts' brokerage houses (Bhushan 1989), which affects analyst forecasting behavior. We include the inverse of stock prices (*INVPRICE*) because Brennan and Hughes (1991) suggest that it proxies for the brokerage commission rate. Analysts are more likely to follow firms with higher levels of return variability because the anticipated trading benefits based on private information on these stocks are greater (Bhushan 1989). We therefore include *STDROE*, which is measured as the standard deviation of *ROE* in the preceding four quarters, and *RETVAR*, computed as the daily stock return variance over the 200 days prior to the year-end. We include research and development expense (*RD*) as a proxy for the level of information asymmetry (Aboody and Lev 2000) because analysts have relatively stronger incentives to follow firms with higher levels of information asymmetry (Barth et al. 2001). The earnings-return (Pearson) correlation (*CORR*) between *ROE* and annual stock returns in the preceding four quarters captures the difficulty in predicting a firm's earnings. In addition, *ROA* controls for firm profitability. Finally, annual change in *EPS* (*ACHEPS*) controls for the magnitude of the forthcoming earnings information (Ali et al. 2007). All other variables are as defined earlier.

Table 6, Panel A presents a comparison of the levels of and changes in the three main analyst variables in the year following first-time CSR disclosures. Initiators are covered by more analysts than non-initiators (*COVERAGE*: 26.08 for initiators versus 15.72 for non-initiators, $p < 0.01$), and achieve greater improvement in forecast accuracy than non-initiators ($\Delta|FE|$: -0.137 for initiators versus 0.120 for non-initiators, though at a more marginal statistical significance level with $p = 0.07$).

We present the multivariate regression results for Equations (5), Equation (6), and Equation (7) in Panels B, C, and D of Table 6, respectively. Column I of Panel B shows that there is a significantly positive coefficient on *DISCI* $*$ *HIPERFORM* (coeff. $= 1.052$, $p = 0.05$), which suggests that analyst following increases for initiators with superior CSR performance. When we run the regression separately in the two subsamples portioned based on industry medians of CSR

TABLE 6

Post-CSR Disclosure Changes in Analyst Forecasts

Panel A: Mean Comparison

	Full Sample	$DISCI_t = 1$	$DISCI_t = 0$	t-value (difference)
$COVERAGE_{t+1}$	15.768	26.079	15.718	7.95
$\|FE\|_{t+1}$	0.226	0.215	0.226	−0.22
$DISP_{t+1}$	0.205	0.237	0.205	1.08
$\Delta COVERAGE_{t+1}$	−1.250	−0.753	−1.252	−1.04
$\Delta\|FE\|_{t+1}$	0.011	−0.137	0.120	−1.85
$\Delta DISP_{t+1}$	0.009	−0.011	0.009	−0.70

Panel B: Post-CSR Disclosure Analyst Coverage (Dependent Variable = $\Delta COVERAGE_{t+1}$)

	I		II		III	
	Full Sample		Top 50% *PERFORM*		Bottom 50% *PERFORM*	
Variables	**Coeff.**	**t-stat.**	**Coeff.**	**t-stat.**	**Coeff.**	**t-stat.**
Intercept	−0.409***	−16.88	−0.401***	−8.94	−0.506***	−21.86
$DISCI_t$	−0.439	−1.01	0.904**	2.12	0.008	0.03
$HIPERFORM_t$	−0.020	−0.50				
$DISCI_t * HIPERFORM_t$	1.052**	1.95				
$\Delta SIZE_t$	0.493***	8.45	0.855***	5.59	0.471***	7.71
$\Delta STDROE_t$	−0.112	−0.93	−0.201	−0.68	−0.076	−0.60
$\Delta INVPRICE_t$	4.201***	8.44	9.552***	6.53	3.229***	6.41
$\Delta RETVAR_t$	0.008	0.83	−0.009	−0.44	0.027***	2.46
ΔRD_t	3.660***	7.89	0.144	0.10	4.056***	8.65
ΔROA_t	1.751***	7.11	2.189***	3.60	1.616***	6.18
$\Delta CORR_t$	0.001	0.89	0.001	0.69	0.001	1.21
Adjusted R^2	0.026		0.016		0.017	

(continued on next page)

Panel B: Post-CSR Disclosure Analyst Coverage (Dependent Variable = $\Delta COVERAGE_{t+1}$)

	I		II		III	
	Full Sample		Top 50% *PERFORM*		Bottom 50% *PERFORM*	
Variables	**Coeff.**	**t-stat.**	**Coeff.**	**t-stat.**	**Coeff.**	**t-stat.**
n ($DISCI_t = 1$)	213		165		48	
n	15,298		7,664		7,634	

Panel C: Post-CSR Disclosure Analyst Forecast Errors (Dependent Variable = $\Delta\|FE\|_{t+1}$)

	I		II		III	
	Full Sample		Top 50% *PERFORM*		Bottom 50% *PERFORM*	
Variables	**Coeff.**	**t-stat.**	**Coeff.**	**t-stat.**	**Coeff.**	**t-stat.**
Intercept	1.134***	76.49	0.923***	54.07	1.124***	66.68
$DISCI_t$	−0.496**	−1.92	−0.251*	−1.77	−0.425	−1.42
$HIPERFORM_t$	−0.261***	−5.41				
$DISCI_t * HIPERFORM_t$	0.318	1.04				
$\Delta SIZE_t$	−0.277***	−10.74	−0.198***	−5.10	−0.293***	−5.56
$\Delta STDROE_t$	0.129*	1.71	0.059	0.52	0.189*	1.89
$\Delta ACHEPS_t$	−0.131***	−2.59	−0.065	−0.85	−0.140**	−2.09
ΔRD_t	−0.410*	−1.66	−1.513***	−3.83	0.250	0.79
ΔROA_t	0.610***	7.46	0.187	1.28	0.855***	8.07
$\Delta CORR_t$	−0.001	−0.29	0.001	0.32	−0.001	−0.36
Adjusted R^2	0.020		0.011		0.016	
n ($DISCI = 1$)	213		165		48	
n	13,186		6,587		6,599	

(continued on next page)

Panel D: Post-CSR Disclosure Analyst Forecast Dispersion (Dependent Variable = $\Delta DISP_{t+1}$)

	I		II		III	
	Full Sample		Top 50% *PERFORM*		Bottom 50% *PERFORM*	
Variables	Coeff.	t-stat.	Coeff.	t-stat.	Coeff.	t-stat.
Intercept	0.007***	5.59	0.006***	4.67	0.007***	4.98
$DISCI_t$	0.008	0.37	−0.048***	−3.21	−0.020	−1.05
$HIPERFORM_t$	0.001	0.24				
$DISCI_t * HIPERFORM_t$	−0.053**	−2.08				
$\Delta SIZE_t$	−0.015***	−6.67	−0.008**	−2.40	−0.021***	−6.62
$\Delta STDROE_t$	0.002	0.33	−0.005	−0.62	0.010	1.09
$\Delta ACHEPS_t$	−0.064***	−6.34	−0.100***	−7.12	−0.031**	−2.11
ΔRD_t	−0.022	−1.04	0.057*	1.76	−0.071**	−2.38
ΔROA_t	−0.051***	−4.10	−0.058***	−3.44	−0.053***	−2.81
$\Delta CORR_t$	−0.001	−0.49	0.001	1.67	−0.001	−2.01
Adjusted R^2	0.024		0.011		0.011	
n ($DISCI = 1$)	213		165		48	
n	14,363		7,153		7,210	

*, **, *** Indicate that the estimated coefficient is statistically significant at the 10 percent, 5 percent and 1 percent levels, respectively.

All continuous variables are winsorized at the 1st and 99th percentiles. All t-statistics are corrected using the Huber-White procedure.

Variable Definitions:

$DISCI_t$ = indicator variable that equals 1 if it is the earliest reporting year of a firm that issues CSR reports (CSR report initiating year), and 0 otherwise;

$HIPERFORM_t$ = indicator variable that equals 1 if the firm's total CSR strength score, namely, *PERFORM*, is higher than the industry median (in other words, if the firm is classified as a high CSR performer in its industry), and 0 otherwise;

$\Delta COVERAGE_{t+1}$ = $COVERAGE_{t+1} - COVERAGE_t$; where *COVERAGE* is the 12-month average of the number of analysts who issued annual earnings forecasts in I/B/E/S for the firm of interest;

$\Delta|FE|_{t+1}$ = $|FE|_{t+1} - |FE|_t$; where $|FE|$ is the absolute value of the 12-month average of analyst forecast errors defined as actual earnings minus the mean forecast, deflated by the stock price at the beginning of the fiscal year;

$\Delta DISP_{t+1}$ = $DISP_{t+1} - DISP_t$; where *DISP* is the 12-month average of the standard deviation of analyst forecasts, deflated by the stock price at the beginning of the fiscal year;

SIZE = natural logarithm of the market value of equity (CSHO * PRCC_F) at the beginning of the year;

(continued on next page)

STDROE = standard deviation of ROE (IB/CEQ) in the four quarters of the preceding year;
INVPRICE = inverse of the stock price (PRCC_F) at the beginning of the fiscal year;
RETVAR = daily stock return variance estimated over the 200 days prior to the year-end;
RD = research and development expense (XRD) deflated by total assets at the beginning of the fiscal year;
ROA = earnings before extraordinary items divided by total assets; and
CORR = Pearson correlation coefficient between *ROE* and annual stock returns in the four quarters of the preceding year.

performance, we find a significantly positive coefficient on *DISCI* (coeff. = 0.904, p = 0.03) but only in the better performing group.[21]

We obtain similar results for the absolute magnitude of analyst forecast errors. Table 6, Panel C demonstrates that, while in the full sample, the coefficient on *DISCI* * *HIPERFORM* is insignificant, the main effect of *DISCI* in the partitioned sample regression is marginally negative (coeff. = −0.251, p = 0.08) for the better performing group. The results for forecast dispersion, presented in Table 6, Panel D show a similar pattern. In the full-sample regression in Column I, the coefficient on *DISCI* * *HIPERFORM* is significantly negative (coeff. = −0.053, p = 0.04). The regressions in the partitioned sample (Columns II and III) yield a significantly negative coefficient on *DISCI* only among better performing firms (coeff. = −0.048, $p < 0.01$).

In sum, voluntary CSR disclosure is associated with increased analyst coverage, improved forecast accuracy, and a reduction in forecast dispersion among firms with relatively superior CSR performance. These results are consistent with our conjecture that CSR disclosure by strong CSR performers helps reduce information asymmetry between managers and shareholders and among shareholders. The evidence supports our reasoning that CSR disclosure can reduce the cost of equity capital by reducing estimation risk in the market. In other words, consistent with our earlier evidence that the cost of equity capital benefit manifests only among firms with relatively superior CSR performance, the effects of CSR disclosure on analyst coverage and forecast accuracy and dispersion are significant only when firms achieve relatively superior CSR performance.

CSR Disclosure and Subsequent Equity Issuances

As discussed previously, we predict that firms anticipating external financing needs are more likely to initiate CSR disclosures in the hope of obtaining cheaper capital. Hence, we should observe more equity issuances after first-time CSR disclosures. We estimate the following logistic regression for Equation (8) and OLS regression for Equation (9) to empirically test this prediction:

$$\begin{aligned}\log[\text{prob}(SEO_{i,t+T})/(1-\text{prob}(SEO_{i,t+T}))] = {}& \beta_0 + \beta_1 DISCI_{i,t} + \beta_2 MB_{i,t} + \beta_3 LNSALES_{i,t} \\ & + \beta_4 ROA_{i,t} + \beta_5 LEV_{i,t} + \beta_6 CASH_{i,t} + \beta_7 FIN_{i,t} \\ & + \beta_8 PAYOUT_{i,t} + \beta_9 CAPITAL_{i,t} + \beta_{10} RD_{i,t} \\ & + \beta_{11} LNDISP_{i,t} + \varepsilon_{i,t}, \end{aligned} \tag{8}$$

$$\begin{aligned} ISSUE\$_{i,t+T} = {}& \beta_0 + \beta_1 DISCI_{i,t} + \beta_2 MB_{i,t} + \beta_3 LNSALES_{i,t} + \beta_4 ROA_{i,t} + \beta_5 LEV_{i,t} + \beta_6 CASH_{i,t} \\ & + \beta_7 FIN_{i,t} + \beta_8 PAYOUT_{i,t} + \beta_9 CAPITAL_{i,t} + \beta_{10} RD_{i,t} + \beta_{11} LNDISP_{i,t} + \varepsilon_{i,t}, \end{aligned} \tag{9}$$

where T (= 1 or 2) denotes one or two years following CSR disclosure, and other notations follow those of earlier regression equations. SEO_{t+1} (SEO_{t+2}) equals 1 if a firm conducts a seasoned equity offering within one (two) year(s) following CSR disclosure. $ISSUE\$_{t+1}$ ($ISSUE\$_{t+2}$) is the total dollar amount in billions raised in SEOs within one (two) year(s) following CSR disclosure. We obtain information on SEOs from the Security Data Corporation (SDC).

Following prior studies, we control for other potential factors affecting the equity issuance decisions of firms. We include the market-to-book ratio (*MB*), as Stein (1995) suggests that firms will choose the time when their stock is overvalued to issue equity. In addition, growth firms have

[21] A careful examination of the distribution of $\Delta COVERAGE$ reveals that there are some relatively large values on the negative tail. To determine whether our results are sensitive to these large values, we exclude the bottom 5 percent of $\Delta COVERAGE$ and get a more symmetric distribution on this variable. We also try excluding extreme values of the upper and the bottom 2 percent of this variable. Overall, our main results are not sensitive to the treatment of the large negative values of $\Delta COVERAGE$.

a greater need for capital. We include research and development expenses (*RD*) and capital expenditures (*CAPITAL*) as additional proxies for growth opportunities. The likelihood of an equity issuance can depend on the extent of the financial constraints that a firm faces. Also, firms may follow a specific pecking order in their choice of financing options and rely preferentially on internal reserves and debt financing before issuing equity (Myers and Majluf 1984). Hence, we include profitability (*ROA*), cash flow (*CASH*), payout ratio (*PAYOUT*: cash dividend) and leverage (*LEV*) to capture financial constraints, internally generated funds, and debt capacity. We control for analyst forecast dispersion (*LNDISP*: the logarithm of the standard deviation of analyst forecasts divided by the consensus forecast) as a proxy for the degree of agreement between management and investors because Dittmar and Thakor (2007) argue that firms are more likely to issue equity when the level of agreement is high. Finally, we control for firm size (*LNSALES*: logarithm of total sales) and financing activities (*FIN*) already conducted in the current year.

Equation (8) assesses whether CSR disclosure is related to the likelihood of future equity issuance through SEOs, and Equation (9) examines whether CSR disclosure is associated with the size of SEOs. Table 7, Panel A indicates that firms are more likely to seek equity capital through SEOs in the two years following CSR disclosure. The coefficients on *DISCI* are positive and marginally significant (coeff. = 0.504, p = 0.09 for SEO_{t+1}; coeff. = 0.625, p = 0.06 for SEO_{t+2}). Based on the coefficient estimate for *DISCI* in Column I (II), in the first year (two years) after CSR disclosure, the odds that initiators will issue equity is 65.5 percent (86.8 percent) higher than that of non-initiators.

Table 7, Panel B reveals that, holding other factors constant, disclosing firms not only are more likely to issue equity, but also raise a significantly larger amount than non-initiators. The difference ranges from U.S. $165 million ($ISSUE\$_{t+2}$) to U.S. $173 million ($ISSUE\$_{t+1}$). These results are consistent with managers initiating CSR disclosures before going to the capital market in anticipation of obtaining cheaper external capital and increasing their capacity to raise external capital.

Additional Analyses

Alternative Measures of CSR Disclosure

Our CSR disclosure measure, *DISCI*, which captures first-time reporters, best serves the purpose of testing our hypotheses. However, for the sake of completeness, we also test our hypotheses using alternative measures of CSR disclosure. The first measure, $DISC_{i,t}$, is an indicator variable that equals 1 if Firm *i* discloses a standalone CSR report in Year *t*, and 0 otherwise. The second measure, $DISCN_{i,t}$, indicates whether a firm only sporadically publishes CSR reports. $DISC_{i,t}$ takes a value of 1 if Firm *i* issues a CSR report in year *t*, but not in year *t*+1 (even though it resumes reporting in year *t*+2 or later), and 0 otherwise. Allowing for non-first-time disclosing years significantly increases our sample size. We obtain results qualitatively similar to those based on *DISCI*. Finally, after excluding first-time reports, we use only continuing reports, $DISC_{i,t}$, which can be just mundane duplications of earlier reports containing less incremental information. We find a much weaker, albeit still significant, result.

Individual Measures of the Cost of Equity Capital

In the above analyses, we use the average of the three cost of equity capital measures based on Gebhardt et al. (2001), Claus and Thomas (2001), and Easton (2004). These estimation methods are based on different earnings growth assumptions and therefore have distinct strengths and weaknesses. The merit of each measure is debated among researchers, and it is not our intention to resolve this debate. Averaging across the three estimates potentially reduces noise in individual measures (Larcker and Rusticus 2010), and is widely used in the literature (Hail and Leuz 2006;

TABLE 7

Post-CSR Disclosure Seasoned Equity Offerings

	Panel A: Incidence of SEOs				Panel B: Size of SEOs			
	Column I		Column II		Column I		Column II	
	SEO_{t+1}		SEO_{t+2}		$ISSUEAMT_{t+1}$		$ISSUEAMT_{t+2}$	
Variables	Coeff.	Prob.	Coeff.	Prob.	Coeff.	t-stat.	Coeff.	t-stat.
Intercept	−1.700***	0.00	−2.826***	0.00	−0.406***	−10.76	−0.380***	−6.65
$DISCI_t$	0.504*	0.09	0.625*	0.06	0.173***	2.95	0.165**	2.32
MB_t	0.024***	0.01	0.013	0.25	0.008***	4.38	0.013***	4.63
$LNSALES_t$	−0.234***	0.00	−0.138***	0.00	0.077***	16.71	0.076***	11.60
ROA_t	−1.475***	0.00	−2.046***	0.00	0.257***	3.09	0.260**	2.02
LEV_t	2.599***	0.00	2.487***	0.00	0.115***	3.55	0.096**	1.95
$CASH_t$	−0.968***	0.00	−0.582*	0.10	0.121**	2.38	0.115	1.46
FIN_t	2.135***	0.00	1.310***	0.00	0.059	1.56	0.018	0.30
$PAYOUT_t$	1.261*	0.10	0.554	0.43	0.202***	2.94	0.069	0.26
$CAPITAL_t$	0.870***	0.00	0.772***	0.00	0.104***	3.71	0.058	1.39
RD_t	−1.891**	0.03	−2.483**	0.03	0.417***	2.78	0.494**	2.06
$LNDISP_t$	0.046	0.73	−0.134	0.44	0.073***	3.05	0.090***	2.63
Pseudo-R^2	0.109		0.073		0.322		0.307	
Likelihood Ratio	567.38		272.51					
Percent Concordant	72.50		70.00					
n ($DISCI = 1$)	189		189					
n ($SEO = 1$)	885		517					
n	13,013		13,013		885		517	

*, **, *** Indicate that the estimated coefficient is statistically significant at the 10 percent, 5 percent and 1 percent levels, respectively.

All continuous variables are winsorized at the 1st and 99th percentiles. All t-statistics are corrected using the Huber-White procedure.

Variable Definitions:

(continued on next page)

TABLE 7 (continued)

$DISCI_t$ = indicator variable that equals 1 if it is the earliest reporting year of a firm that issues CSR reports (CSR report initiating year), and 0 otherwise;
$HIPERFORM_t$ = indicator variable that equals 1 if the firm's total CSR strength score, namely, *PERFORM*, is higher than the industry median (in other words, if the firm is classified as a high CSR performer in its industry), and 0 otherwise;
SEO_{t+1} (SEO_{t+2}) = 1 if a firm conducts a seasoned equity offering within one (two) year(s) after CSR disclosure, and 0 otherwise;
$ISSUEAMT_{t+1}$ ($ISSUEAMT_{t+2}$) = total dollar amount in billions issued by a firm's seasoned equity offering within one (two) year(s) after the firm issues a CSR report;
MB_t = market-to-book ratio, which is defined as the market value of equity (CSHO * PRCC_F) divided by the book value of equity (CEQ);
$LNSALES_t$ = natural logarithm of total sales (SALE);
LEV_t = leverage ratio, which is defined as the ratio of total debt (DLTT + DLC) divided by total assets;
$CASH_t$ = total cash-to-asset ratio, which is defined as cash and short-term investments (CHE) divided by total assets;
FIN_t = amount of debt or equity capital raised by the firm scaled by total assets at the beginning of year *t*. It is measured as the sale of common stock and preferred shares minus the purchase of common stock and preferred shares (SSTK – PRSTKC) plus the long-term debt issuance minus the long-term debt reduction (DLTIS – DLTR);
$PAYOUT_t$ = cash dividend (DVC) relative to total assets;
$CAPITAL_t$ = capital expenditures (CAPX) scaled by total assets;
RD_t = research and development expense (XRD) scaled by total assets. We assume a value of 0 if a firm's research and development expense is missing; and
$LNDISP_t$ = forecast dispersion in year *t*, which is measured as the logarithm of the standard deviation of analyst estimates of year *t* earnings divided by the consensus forecast of year *t* earnings.

Dhaliwal et al. 2006). To provide assurance that our results are not sensitive to the choice of these measures, we repeat our analyses using the three measures separately and obtain similar results.

Correlation between KLD Performance and CSR Disclosure

KLD performance scores (*PERFORM*) are correlated with the decision of firms to issue CSR reports for the first time (*DISCI*), but their correlation coefficients (Table 2, Panel B) are moderate at about 10 percent.[22] As discussed earlier, we control for CSR performance in all regression equations. To further alleviate the concern regarding the correlation between the KLD performance scores and CSR disclosure of firms and the potential impact of this correlation on our results, we conduct the following robustness analyses. First, we remove the transparency-related category, that is, "Corporate Governance," from the KLD performance ranking scores, as this category contains a subcategory, "Transparency," which is a dimension that is likely to reflect the CSR disclosure policy of firms. Our main inferences are unchanged. Second, we use the performance score of each KLD CSR category to measure firms' social performance. Our main inferences are unchanged. Finally, we match each *DISCI* observation with a non-disclosing firm that has the closest industry-adjusted KLD CSR performance score in the same year and industry and run regression Equation (2). The coefficient on *DISCI* is significantly negative at the conventional level, suggesting that CSR initiators enjoy a subsequent reduction in the cost of equity capital.

Alternative Measures of CSR Performance

To determine if our results are sensitive to alternative measures of CSR performance, we use two additional measures. One measure is an indicator *DJSI* that equals 1 if a firm appeared in the Dow Jones Sustainability Index in any year during the period 2002–2007, and 0 otherwise. The other is an indicator *CRO* that equals 1 if a firm was on the "100 Best Corporate Citizens" list for 2007 from the Corporate Responsibility Officer, and 0 otherwise.[23] We do not consider year-to-year variation because of data constraint. It turns out *DJSI* (*CRO*) is correlated with our KLD performance scores with a Pearson coefficient of 30 percent (23 percent). Using these two measures in place of the KLD scores produces qualitatively similar results.

V. SUMMARY AND CONCLUSIONS

We examine a potential benefit associated with the initiation of voluntary disclosure of CSR activities: a reduction in the cost of equity capital. We find that the likelihood of a firm initiating standalone disclosure of CSR activities is associated with a higher prior year cost of equity capital. Firms with CSR performance superior to that of their industry peers enjoy a reduction in the cost of equity capital after they initiate CSR reports. Further, firms initiating CSR disclosure with superior CSR performance attract dedicated institutional investors and analyst coverage, and these analysts achieve lower absolute forecast errors and dispersion following such disclosure. Finally, CSR disclosure initiators appear to exploit this potential benefit of a reduction in the cost of equity

---

[22] Consistent with this low correlation, KLD provides the following information regarding how it rates the CSR performance of firms (see http://www.kld.com/research/methodology.html): "KLD researches the social, environmental, and governance performance of corporations. KLD research relies on five distinct data sources to inform our ratings and analysis. Data are collected in a disciplined process from a wide variety of company, government, and non-government organization and media sources. KLD tracks each company through more than 14,000 global media sources daily." These five distinct data sources include (1) direct communication with company officers; (2) a network of global ESG research firms that cover non-U.S. markets; (3) review of more than 14,000 global news sources; (4) public documents of companies, including annual reports and proxy statements; and (5) information obtained from government and non-government organizations including the U.S. Department of Labor, EPA, Human Rights Watch, OSHA, CANICOR, Ceres, ICCR, and DoD. Hence, it appears that the CSR reports of firms constitute only one of the numerous information sources employed by KLD.

[23] See http://www.thecro.com/.

capital. They are more likely than non-disclosing firms to conduct SEOs to raise capital in the two years following the disclosure. In addition, among firms conducting SEOs, CSR disclosure initiators raise a significantly larger amount of equity capital than non-initiators.

This study adds to the voluntary disclosure literature by extending the traditional research on voluntary disclosure beyond the narrow focus of financial disclosure. Our analyses enhance our understanding of the rationales behind and the consequences of the recent trend in voluntary CSR disclosure. These results have important implications for companies, regulators, and investors.

A few caveats are worth noting. Most of the control variables that we use in the CSR determination model are obtained from the standard voluntary disclosure literature. To the extent that CSR disclosure is distinct from other forms of voluntary disclosure examined in the literature, we may have missed important determinants of CSR disclosure. In addition, it is possible that we missed some reports on stale websites because of their lack of maintenance, which would add noise to our results. Also, we do not examine the content of the CSR reports. To the extent that the detailed information of these reports is not fully captured by the KLD scores, we fail to capture some important characteristics of CSR reports. Further, it is important to control for the other disclosure policies of firms when examining the impact of CSR disclosure. Our empirical proxies using management guidance and earnings quality may not be sufficient to capture these potential confounding effects. Finally, although the KLD rating is widely used in the management literature, a significant amount of future research is warranted to further establish its validity in measuring the social performance of firms.

These caveats notwithstanding, we believe that our study opens various venues for future research. For example, CSR disclosure and performance could have a different impact on the cost of debt as debtholders have a payoff function different from that of equityholders. Further, the effect of CSR disclosure could be a function of differences in legal environment and institutional setting. Therefore, an international study could help us better understand CSR disclosure. Last, as mentioned previously, it would be worthwhile to investigate the effect of the information content of CSR reports on the valuation decisions of investors.

APPENDIX
CSR CATEGORIES AND PERFORMANCE SCORES BY INDUSTRY

Industries	Total Strength	Community	Corporate Governance	Diversity	Employee Relations	Environment	Human Rights	Product
Mining/Construction	0.789	0.143	0.130	0.189	0.200	0.084	0.027	0.016
Food	2.351	0.399	0.231	1.065	0.401	0.203	0.022	0.028
Textiles/Print/Publish	1.751	0.159	0.143	0.670	0.322	0.296	0.003	0.158
Chemicals	2.224	0.298	0.145	0.713	0.525	0.442	0.000	0.101
Pharmaceuticals	1.599	0.201	0.136	0.752	0.309	0.116	0.000	0.085
Extractive	1.266	0.147	0.159	0.210	0.488	0.241	0.002	0.020
Manf: Rubber/Glass/etc.	1.638	0.311	0.168	0.485	0.404	0.162	0.036	0.072
Manf: Metal	1.186	0.060	0.126	0.204	0.426	0.285	0.000	0.085
Manf: Machinery	1.300	0.134	0.128	0.316	0.318	0.237	0.004	0.163
Manf: Electrical Eqpt	1.397	0.138	0.135	0.426	0.331	0.182	0.007	0.179
Manf: Transport Eqpt	1.985	0.193	0.176	0.606	0.537	0.330	0.000	0.144
Manf: Instruments	1.398	0.127	0.184	0.557	0.230	0.174	0.001	0.126
Manf: Misc.	1.970	0.463	0.232	0.713	0.335	0.055	0.030	0.140
Computers	1.687	0.120	0.183	0.769	0.391	0.114	0.001	0.110
Transportation	1.273	0.118	0.139	0.631	0.269	0.041	0.002	0.076
Utilities	1.919	0.203	0.173	0.721	0.364	0.440	0.001	0.016
Retail: Wholesale	0.757	0.030	0.119	0.341	0.122	0.065	0.000	0.081
Retail: Misc.	1.598	0.192	0.149	0.870	0.230	0.070	0.016	0.071
Retail: Restaurant	1.545	0.110	0.140	0.932	0.235	0.110	0.004	0.015
Financial	1.624	0.352	0.265	0.710	0.227	0.009	0.005	0.055
Insurance/Real Estate	0.534	0.036	0.153	0.256	0.085	0.004	0.000	0.001
Services	0.822	0.026	0.146	0.492	0.106	0.020	0.000	0.032
Mean	1.478	0.180	0.174	0.611	0.292	0.137	0.005	0.078

This table provides a brief summary of the seven categories included in the KLD Research & Analytics database, which are used to rate the CSR performance of firms. The data cover the 1991–2007 period. Within each of these seven categories, KLD defines a set of potential strengths and assigns a value of 1 if the strength exists, and 0 otherwise. The statistics provided in this table are the mean performance scores (non-industry-adjusted raw performance scores) for each industry. The seven main categories and their subcategories are summarized below.

Main Categories	Sub-Categories	Max. Strength (Perfect Score)	Actual Max. Strength	Actual Mean Strength
Community	(1) Charitable Giving, (2) Innovative Giving, (3) Non-U.S. Charitable Giving, (4) Support for Education, (5) Support for Housing, (6) Volunteer Programs, and (7) Other Strengths	7	5	0.189
Corporate Governance	(1) Compensation, (2) Ownership, (3) Political Accountability, (4) Transparency, and (5) Other Strengths	5	3	0.167
Diversity	(1) Board of Directors, (2) CEO, (3) Employment of the Disabled, (4) Promotion, (5) Women and Minority Contracting, (6) Work/Life Benefits, (7) Gay and Lesbian Policies, and (8) Other Strengths	8	7	0.605
Employee Relations	(1) Health and Safety, (2) Retirement Benefits, (3) Union Relations, (4) Cash Profit Sharing, (5) Employee Involvement, and (6) Other Strengths	6	5	0.292
Environment	(1) Beneficial Products and Services, (2) Clean Energy, (3) Pollution Prevention, (4) Recycling, and (5) Other Strengths	5	4	0.140
Human Rights	(1) Labor Rights, (2) Relations with Indigenous Peoples, and (3) Other Strengths	3	2	0.004
Product	(1) Benefits the Economically Disadvantaged, (2) Quality, (3) R&D/Innovation, and (4) Other Strengths	4	3	0.077
Total Strength	The sum of all of the above seven main categories.	38	29	1.474

REFERENCES

Aboody, D., and B. Lev. 2000. Information asymmetry, R&D, and insider gains. *The Journal of Finance* 55: 2747–2766.

Ali, A., T. Y. Chen, and S. Radhakrishnan. 2007. Corporate disclosures by family firms. *Journal of Accounting and Economics* 44 (1–2): 238–286.

Al-Tuwaijri, S. A., T. E. Christensen, and K. E. Hughes, II. 2004. The relations among environmental disclosure, environmental performance, and economic performance: A simultaneous equations approach. *Accounting, Organizations and Society* 29: 447–471.

Amihud, Y., and H. Mendelson. 1986. Asset pricing and the bid-ask spread. *Journal of Financial Economics*

17: 223–249.
Anderson, J. C., and A. W. Frankel. 1980. Voluntary social reporting: An ISO-beta portfolio analysis. *The Accounting Review* 15: 467–479.
Baron, D. P., M. A. Harjoto, and H. Jo. 2009. The economics and politics of corporate social performance. Working paper, Stanford University.
Barth, M. E., W. H. Beaver, and W. R. Landsman. 1998. Relative valuation roles of equity book value and net income as a function of financial health. *Journal of Accounting and Economics* 25: 1–34.
———, R. Kasznik, and M. F. McNichols. 2001. Analyst coverage and intangible assets. *Journal of Accounting Research* 39: 1–34.
Berman, S. L., A. C. Wicks, S. Kotha, and T. M. Jones. 1999. Does stakeholder orientation matter? The relationship between stakeholder management models and firm financial performance. *Academy of Management Journal* 42: 488–506.
Bhushan, R. 1989. Firm characteristics and analyst following. *Journal of Accounting and Economics* 11: 255–274.
Botosan, C. A. 1997. Disclosure level and the cost of equity capital. *The Accounting Review* 72: 323–349.
———, and M. A. Plumlee. 2002. A re-examination of disclosure level and the expected cost of equity capital. *Journal of Accounting Research* 40: 21–40.
Brennan, M., and P. Hughes. 1991. Stock prices and the supply of information. *The Journal of Finance* 46: 1665–1691.
Bushee, B. J. 1998. The influence of institutional investors on myopic R&D investment behavior. *The Accounting Review* 73: 305–333.
——— 2001. Do institutional investors prefer near-term earnings over long-run value? *Contemporary Accounting Research* 18: 207–246.
———, and C. F. Noe. 2000. Corporate disclosure practices, institutional investors, and stock return volatility. *Journal of Accounting Research* 38: 171–202.
Chatterji, A. K., D. I. Levine, and M. W. Toffel. 2009. How well do social ratings actually measure corporate social responsibility? http://onlinelibrary.wiley.com/doi/10.1111/j.1530-9134.2009.00210.x/abstract 18: 125–169.
Claus, J., and J. Thomas. 2001. Equity premia as low as three percent? Evidence from analysts' earnings forecasts for domestic and international stock markets. *The Journal of Finance* 56: 1629–1666.
Core, J. E. 2001. A review of the empirical disclosure literature: Discussion. *Journal of Accounting and Economics* 31: 441–456.
Cormier, D., and M. Magnan. 2003. Does disclosure matter? *CA Magazine* 136 (4): 43–45.
Dechow, P. M., R. G. Sloan, and A. P. Sweeney. 1995. Detecting earnings management. *The Accounting Review* 70: 193–226.
———, S. A. Richardson, and I. Tuna. 2003. Why are earnings kinky? An examination of the earnings management explanation. *Review of Accounting Studies* 8: 355–384.
Del Guercio, D. 1996. The distorting effect of the prudent-man laws of institutional equity investments. *Journal of Financial Economics* 40: 31–62.
Dhaliwal, D. S., S. Heitzman, and O. Z. Li. 2006. Taxes, leverage, and the cost of equity capital. *Journal of Accounting Research* 44: 691–723.
———, L. Krull, O. Z. Li, and W. Moser. 2005. Dividend taxes and implied cost of equity capital. *Journal of Accounting Research* 43: 675–708.
Diamond, D., and R. Verrecchia. 1991. Disclosure, liquidity and the cost of equity capital. *The Journal of Finance* 46: 1325–1360.
Dietrich, J. R., S. J. Kachelmeier, D. N. Kleinmuntz, and T. J. Linsmeier. 2001. Market efficiency, bounded rationality, and supplemental business reporting disclosures. *Journal of Accounting Research* 39: 243–268.
Dittmar, A., and A. Thakor. 2007. Why do firms issue equity? *The Journal of Finance* 62: 1–54.
Dye, R. A. 1985. Disclosure of non-proprietary information. *Journal of Accounting Research* 23: 123–145.
Easton, P. 2004. PE ratios, PEG ratios, and estimating the implied expected rate of return on equity capital. *The Accounting Review* 79: 73–95.
Fama, E., and K. French. 1992. The cross-section of expected stock returns. *The Journal of Finance* 47:

427–465.
Francis, J., D. Nanda, and P. Olsson. 2008. Voluntary disclosure, earnings quality, and cost of Capital. *Journal of Accounting Research* 46: 53–99.
Francis, J., D. Philbrick, and K. Schipper. 1994. Shareholder litigation and corporate disclosures. *Journal of Accounting Research* 32: 137–164.
Frankel, R., M. McNichols, and P. Wilson. 1995. Discretionary disclosure and external financing. *The Accounting Review* 70: 135–150.
Gebhardt, W., C. Lee, and B. Swaminathan. 2001. Toward an implied cost of capital. *Journal of Accounting Research* 39: 135–176.
Gode, D., and P. Mohanram. 2003. Inferring the cost of capital using the Ohlson-Juettner model. *Review of Accounting Studies* 8: 399–431.
Graham, J. R., C. R. Harvey, and S. Rajgopal. 2005. The economic implications of corporate financial reporting. *Journal of Accounting and Economics* 40: 3–73.
Graves, S. B., and S. A. Waddock. 1994. Institutional owners and corporate social performance. *Academy of Management Journal* 37: 1035–1046.
Hail, L., and C. Leuz. 2006. International differences in cost of equity: Do legal institutions and securities regulation matter? *Journal of Accounting Research* 44: 485–531.
Healy, P., and K. Palepu. 2001. Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. *Journal of Accounting and Economics* 31: 405–440.
Hobson, J. L., and S. J. Kachelmeier. 2005. Strategic disclosure of risky prospects: A laboratory experiment. *The Accounting Review* 80: 825–846.
Hughes, J. S., J. Liu, and J. Liu. 2007. Information asymmetry, diversification, and cost of capital. *The Accounting Review* 82: 705–730.
Ingram, R. W., and K. B. Frazier. 1980. Environmental performance and corporate disclosure. *Journal of Accounting Research* 18: 614–622.
Jones, J. 1991. Earnings management during import relief investigations. *Journal of Accounting Research* 29: 193–228.
KPMG. 2008. KPMG international survey of corporate social responsibility reporting 2008. Available at: http://www.kpmg.com/LU/en/IssuesAndInsights/Articlespublications/Pages/KPMGInternationalSurveyonCorporateResponsibilityReporting2008.aspx.
Lambert, R., C. Leuz, and R. E. Verrecchia. 2007. Accounting information, disclosure, and the cost of capital. *Journal of Accounting Research* 45: 385–420.
Lang, M., and R. Lundholm. 1993. Cross-sectional determinants of analyst ratings of corporate disclosures. *Journal of Accounting Research* 31: 246–271.
———, and ———. 1996. Corporate disclosure policy and analyst behavior. *The Accounting Review* 71: 467–492.
Larcker, D. F., and T. O. Rusticus. 2010. On the use of instrumental variables in accounting research. *Journal of Accounting and Economics* 49 (3): 186–205.
Leftwich, R. W., R. L. Watts, and J. L. Zimmerman. 1981. Voluntary corporate disclosure: The case of interim reporting. *Journal of Accounting Research* 19: 50–77.
Leuz, C., and C. Schrand. 2008. Voluntary disclosure and the cost of capital: Evidence from firms' responses to the Enron shock. Working paper, University of Chicago.
———, and R. Verrecchia. 2000. The economic consequences of increased disclosure. *Journal of Accounting Research* 38: 91–124.
———, and P. Wysocki. 2008. Economic consequences of financial reporting and disclosure regulation: A review and suggestions for future research. Working paper, University of Chicago.
Lev, B., C. Petrovits, and S. Radhakrishnan. 2010. Is doing good good for you? How corporate charitable contributions enhance revenue growth. *Strategic Management Journal* 31 (2): 182–200.
Margolis, J. D., and J. Walsh. 2001. *People and Profits? The Search for a Link between a Firm's Social and Financial Performance*. Mahwah, NJ: Lawrence Erlbaum Publishers.
Matsumoto, D. A. 2002. Management's incentives to avoid negative earnings surprises. *The Accounting Review* 77: 483–514.
McWilliams, A., and D. Siegel. 2001. Corporate social responsibility: A theory of the firm perspective.

Academy of Management Review 26: 117–127.

Merton, R. C. 1987. A simple model of capital market equilibrium with incomplete information. *The Journal of Finance* 42: 483–510.

Myers, S. C., and N. S. Majluf. 1984. Corporate financing and investment decisions when firms have information that investors do not have. *Journal of Financial Economics* 13: 187–221.

Orlitzky, M., F. Schmidt, and S. Rynes. 2003. Corporate social and financial performance: A meta-analysis. *Organization Studies* 24: 403–441.

Plumlee, M., D. Brown, and S. Marshall. 2008. The impact of voluntary environmental disclosure quality on firm value. Working paper, University of Utah.

Richardson, S., S. H. Teoh, and P. Wysocki. 2004. The walkdown to beatable analysts' forecasts: The role of equity issuance and inside trading incentives. *Contemporary Accounting Research* 21: 885–924.

Richardson, A., and M. Welker. 2001. Social disclosure, financial disclosure and the cost of equity capital. *Accounting, Organizations and Society* 26: 597–616.

Rodriguez, P., D. S. Siegel, A. Hillman, and L. Eden. 2006. Three lenses on the multinational enterprise: Politics, corruption and corporate social responsibility. *Journal of International Business Studies* 37: 733–746.

Rogers, J., and A. Van Buskirk. 2009. Shareholder litigation and changes in disclosure behavior. *Journal of Accounting and Economics* 47 (1–2): 136–156.

Shleifer, A., and R. W. Vishny. 1986. Large shareholders and corporate control. *The Journal of Political Economy* 94: 461–488.

Simnett, R., A. Vanstraelen, and W. F. Chua. 2009. Assurance on sustainability reports: An international comparison. *The Accounting Review* 84: 937–967.

Skinner, D. J. 1997. Earnings disclosures and stockholder lawsuits. *Journal of Accounting and Economics* 23: 249–282.

Sletten, E. 2008. The effect of stock price on discretionary disclosure. Working paper, Massachusetts Institute of Technology.

Social Investment Forum. 2007. 2007 report on socially responsible investing trends in the United States. Washington, D.C.: Social Investment Forum.

Stein, J. C. 1995. Rational capital budgeting in an irrational world. Working paper, Massachusetts Institute of Technology.

Verrecchia, R. E. 2001. Essays on disclosure. *Journal of Accounting and Economics* 32: 97–180.

Waddock, S. 2003. Myths and realities of social investing. *Organization & Environment* 16: 369–380.

THE ACCOUNTING REVIEW
Vol. 86, No. 1
2011
pp. 101–130

American Accounting Association
DOI: 10.2308/accr.00000006

Should Investors Follow the Prophets or the Bears? Evidence on the Use of Public Information by Analysts and Short Sellers

Michael S. Drake
The Ohio State University

Lynn Rees
Edward P. Swanson
Texas A&M University

ABSTRACT: We investigate whether short sellers and analysts differ in their use of information that is predictive of future returns. We find that short interest is significantly associated in the expected direction with all 11 variables examined. In contrast, analysts tend to positively recommend stocks with high growth, high accruals, and low book-to-market ratios, despite these variables having a negative association with future returns. We then investigate the profitability of using short interest in trading. We find abnormal returns (1.11 percent per month) from a zero-investment strategy that (1) shorts firms with highly favorable analyst recommendations (buy signal) but high short interest (sell signal), and (2) buys firms with highly unfavorable analyst recommendations (sell signal) but low short interest (buy signal). Short interest, therefore, appears to capture predictive information that can be used by investors in trading against analysts' recommendations to increase returns.

Keywords: *short interest; analyst recommendations; fundamental analysis; arbitrage.*

Data Availability: *Data are available from the sources identified in the study.*

I. INTRODUCTION

Academic research provides extensive evidence that fundamental analysis can be used to earn abnormal returns (e.g., Frankel and Lee 1998; Piotroski 2000; Swanson et al. 2003). This evidence suggests there is a delay in the price discovery process, which has spawned research into the roles that financial analysts play in this delay (Bradshaw et al. 2001; Jegadeesh

The authors gratefully acknowledge the helpful comments of Steven J. Kachelmeier (senior editor) and two anonymous referees, Larry Brown, Ted Christensen, Darren Henderson, Devin Shanthikumar, Kay Stice, David Wood, and workshop participants at Brigham Young University, Texas A&M University, The University of Texas at Arlington, 2009 FARS Midyear Meeting, 2009 AAA Annual Meeting, and Sanford Bernstein 2009 Conference on Controversies in Quantitative Finance.

Editor's note: Accepted by Steven Kachelmeier

Submitted: July 2009
Accepted: May 2010
Published Online: January 2011

et al. 2004, hereafter JKKL; Bradshaw 2004; Barniv et al. 2009). Analysts' recommendations are the end-product from an extensive analysis of information, and they affect market prices by stating a specific course of action that an investor should take (Asquith et al. 2005a; Malmendier and Shanthikumar 2007). JKKL 2004 provide evidence, however, that analysts' recommendations are positively associated with some accounting, valuation, and growth characteristics that have a *negative* association with future returns (e.g., low book-to-market, high sales growth, and high accruals). Analysts' incentives to obtain investment banking business and to generate trading commissions are potential explanations for why they tend to over-recommend these stocks (Lin and McNichols 1998; Barber et al. 2007) and, indirectly, why analysts might contribute to a delay in price discovery.

Our study investigates how short sellers use publicly available information. Short sellers are regarded as particularly sophisticated investors under financial economic theory.[1] Similar to analysts, short sellers invest considerable time and resources in analyzing companies, but they face potentially different incentives. Because short sellers place their own capital at risk, they have strong incentives to fully use predictive information. Research on short sellers' use of fundamental information is limited, but Dechow et al. (2001) find that short sellers use fundamentals-to-valuation ratios to identify stocks that are expected to realize negative future returns. We extend Dechow et al. (2001) by investigating how short sellers utilize 11 items of fundamental information identified by JKKL that are predictive of future returns. We also conduct the JKKL analysis of analyst recommendations for our sample to allow a direct comparison of information use by analysts and short sellers. Our investigation provides further evidence on the role of short sellers in capital markets and their ability to interpret publicly available information. While financial analysts are acknowledged for their information intermediary role, we investigate the possibility that short sellers can also serve as an information intermediary and thereby facilitate the price discovery process.

Following the JKKL analysis of analysts' information use, we rank firms into quintile portfolios based on the consensus analyst recommendation. In a corresponding manner, we rank firms into quintile portfolios by the level of short interest to reflect short seller beliefs about future returns. We then employ ordered logistic regression to examine how analysts and short sellers use the company-specific predictive information. We categorize the information into four general types: accounting, valuation, growth, and momentum. We find that short interest is strongly associated with all 11 information items in the direction consistent with their empirical relation to future returns. This result indicates that short sellers are highly informed about how company-specific information is likely to affect future returns. In contrast, we confirm results in JKKL that analysts tend to positively recommend stocks with high growth, high accruals, and low book-to-market ratios, despite these variables having a negative association with future returns. When we examine recommendation revisions, we find similar evidence.

Our finding that short sellers interpret information properly as it relates to future returns supports an information intermediary role for short sellers. Consistent with this role, the SEC (1999, 3) observes that short sellers can "add to stock pricing efficiency because their transactions *inform the market* of their evaluation of future stock price performance" (emphasis added). Pownall and Simko (2005) provide evidence that short sellers can serve as information intermediaries when analyst following for a firm is low. In this situation, there are limited alternative

---

[1] Diamond and Verrecchia (1987) argue that only informed traders with strong beliefs that stock prices will fall in the near-term will choose to sell stock short. Their reasoning is based on the notion that the high costs of short selling drives out uninformed traders, so that open short positions reflect trades by more informed investors. Boehmer et al. (2008, 491) comment that short sellers "occupy an exalted place in the pantheon of investors as rational, informed market participants who act to keep prices in line."

sources of guidance. Our tests of information use suggest that short sellers can fill a complementary information intermediary role even when coverage by analysts is extensive.[2]

To provide further evidence on short sellers' effectiveness as information intermediaries, we test whether short interest provides value-relevant information about future returns *beyond* that provided by analyst recommendations and the 11 predictive variables. Short positions could provide incremental information because short sellers have sources of information not considered in our models or because they adjust the weights they place on items of information as market conditions change. In a regression model explaining future six-month returns adjusted for characteristic-based portfolio returns (Daniel et al. 1997), we find that the coefficient on short interest is negative (as predicted) and statistically significant after controlling for the information in analyst recommendations and the 11 predictive variables. These results indicate that short interest provides incremental information about future abnormal returns that is orthogonal to the information provided by analysts and the 11 predictive variables. In addition, the significant negative coefficient on the consensus analyst recommendation indicates that more favorable recommendations result in *lower* future returns after controlling for open short interest and the 11 predictive variables. If we consider recommendation revisions, then its coefficient is insignificant and the coefficient on short interest remains negative and significant.

Our analyses and the evidence from JKKL indicate that analysts sometimes provide favorable (unfavorable) recommendations for stocks with characteristics that are associated with negative (positive) future equity returns. Since analyst recommendations influence trading decisions and stock prices (Asquith et al. 2005a; Malmendier and Shanthikumar 2007), their recommendations can provide support for stock prices that have temporarily deviated from their fundamental values. We investigate if short interest can be used to identify such stocks. That is, we examine if investors can use short interest together with analyst recommendations to construct a portfolio that is likely to earn abnormal future returns. To our knowledge, ours is the first study to link these two investment signals—both from highly regarded capital market participants—to forecast future long-run returns.

We use a large sample of monthly observations over the period 1994 to 2006 to test alternative trading strategies. We first construct quintile portfolios based on the consensus analyst recommendation and then examine abnormal returns from a trading strategy that invests long (short) in firms comprising the best (worst) recommendation quintile. We find that the hedge return is modestly *negative* at −26 basis points per month for our test period. This result is perhaps surprising given the high esteem placed on financial analysts within the financial community, but it is consistent with our results on information use and the growing literature that questions the investment value of analyst recommendations (Demirakos et al. 2004; Bradshaw 2004; Barniv et al. 2009). We employ the same sorting procedure to form investment portfolios based on levels of short interest, and we obtain a statistically significant hedge return of 56 basis points per month from selling short firms in the highest short interest quintile and buying firms in the lowest quintile. Interestingly, the monthly returns from buying stocks with low short interest (29 basis points) are similar to the returns from selling short stocks with high short interest (27 basis points).

---

[2] Note that analysts also provide earnings forecasts, price targets, and narrative discussion that can be informative to investors—possibly more informative than their recommendations.

Concurrent research by Boehmer et al. (2010) also finds that significant abnormal returns can be earned by buying stocks with low short interest.[3]

We next examine whether abnormal returns can be improved by using information from both analysts and short sellers. Here, we intersect the analyst recommendation and short interest quintiles to produce 25 portfolios (5 × 5) formed using information from both signals. We find that monthly abnormal returns are insignificant for portfolios containing stocks about which analysts and short sellers strongly concur (e.g., least favorable recommendation and high short interest, or most favorable recommendation and low short interest). In contrast, returns are highly significant for portfolios of stocks in which they strongly conflict, if we trade consistent with the short sellers. Specifically, we find that an investor would obtain an average monthly abnormal return of 111 basis points from a zero-investment strategy that (1) invests long in firms with the worst recommendations (sell signal) but the lowest short interest levels (buy signal), and (2) invests short in firms with the best recommendations (buy signal) but the highest short interest levels (sell signal). The monthly abnormal return from this strategy is statistically significant in each sub-period, ranging from 71 basis points in 2004–2006 to 130 basis points in 1999–2003. Our dual-signal approach, therefore, provides the most investment value during the volatile 1999–2003 sub-period. By comparison, following analyst recommendations would cause investors to experience sizable losses over this sub-period, and the returns to trading on short interest in isolation would be statistically insignificant.

Our study contributes to our understanding of short sellers by (1) documenting their efficient use of a number of predictive variables discussed in the academic literature, (2) showing that open short positions are incrementally useful in predicting future returns after controlling for those predictive variables, and (3) showing that the returns to mimicking the trading of short sellers are much larger when conditioned on conflicting analyst recommendations. Collectively, these findings provide a more complete picture of how short sellers influence equity price formation and should be of interest to academics, investors, and regulators. This topic has assumed considerable importance due to the alleged role of short selling in the dramatic decline in stock prices that began with the 2008 credit crisis (Boehmer et al. 2009). Academics are also likely to be interested in the implications of our empirical results for Miller's (1977) theory that binding short sale constraints cause pessimists to be under-represented in price formation, leading to overvaluation when a strong divergence of opinion exists about a stock. Our evidence shows that short sellers are under-represented in price formation whenever they disagree with analysts, regardless of whether they are the optimists or the pessimists.

Section II describes the selection of our sample. In Section III, we investigate the relation of short interest and analyst recommendations with information that has been shown by prior research to predict future returns. Section IV presents returns from trading strategies that use analyst recommendations, short interest, or both of these signals together. This section also reports the results of several robustness tests. We discuss our results and conclude in Section V.

II. SAMPLE SELECTION AND DESCRIPTIVE STATISTICS

To perform our analysis, we require time-series data on analyst recommendations, open short positions, the 11 predictor variables, and stock returns. We obtain analyst recommendations from

[3] This finding complements earlier studies that find that stocks with high short interest have significant negative abnormal returns (e.g., Asquith and Meulbroek 1995; Asquith et al. 2005b; Desai et al. 2002). Notably, abnormal returns from trading on high short interest are not necessarily indicative of market inefficiency due to limits to arbitrage. In contrast, the finding that stocks with low short interest have significant positive returns is clearly inconsistent with market efficiency since buy-and-hold strategies are not subject to similar constraints (see Boehmer et al. [2010] for further discussion).

the Thompson Financial I/B/E/S Recommendations database. Beginning in late 1993, I/B/E/S provides analyst recommendations for a wide cross-section of firms. I/B/E/S codes recommendations into five ordered categories: strong buy = 1; buy = 2; hold = 3; sell = 4; and strong sell = 5. For analyses using recommendations, we reverse this coding (i.e., strong buy = 5; strong sell = 1) to allow for a more intuitive interpretation of our results. Each month, we calculate the consensus recommendation level (*Rec*) as the mean of all outstanding recommendations issued a maximum of 12 months prior to month-end.[4] We use only the most recent individual analyst recommendation in the calculation. We also require that the individual analyst recommendations be issued on or before the I/B/E/S consensus recommendation date.[5]

In the first set of analyses, we conduct separate tests of analysts' and short sellers' use of information that is predictive of future returns. We perform these tests using analyst recommendations and short interest as of the last month in each calendar quarter, consistent with JKKL.[6] We include recommendation revisions in these analyses, given that prior research finds that recommendation revisions might be better indicators of future stock price performance than recommendation levels (Womack 1996; JKKL; Barber et al. 2010). We calculate recommendation revisions as the change in recommendation levels from calendar quarter $t-1$ to quarter t (i.e., consecutive quarters). An increase (decrease) in the consensus recommendation indicates an upgrade (downgrade) in the stock relative to the previous calendar quarter $t-1$.[7]

We obtain short interest data to correspond with our analyst recommendations sample period. The Compustat Monthly Securities Database contains monthly short interest for all firms listed on U.S. exchanges beginning in 2003. For earlier years, we purchased monthly short interest data directly from the NYSE, AMEX, and NASDAQ exchanges, and from an online independent vendor.[8] The stock exchanges report open short positions using the 15th of each calendar month as the settlement date (or the last business day before the 15th). We scale short interest by the number of shares outstanding as reported by CRSP and label the resulting ratio as *SIratio*, which is standard in the literature (e.g., Dechow et al. 2001; Asquith et al. 2005b; Pownall and Simko 2005).[9]

---

4 This requirement helps alleviate concerns that our results are being unduly influenced by stale recommendations, and it is similar to the measure used by JKKL. Thompson Financial claims that recommendations not updated for 180 days are excluded from the I/B/E/S consensus recommendation (see Thompson Financial 2009,11); however, we are uncertain as to how long this policy has been in place. Given our long sample period, we follow procedures implemented in prior research.

5 More specifically, I/B/E/S calculates the consensus recommendation on the Thursday before the third Friday of every month (ranging from the 14th to the 20th day of the month). The requirement of excluding recommendations that are issued after this date results in an average delay of 13.7 days between the time when the consensus recommendation is calculated and the beginning of the returns accumulation period. This methodology serves two purposes. First, short interest data are made publicly available mid-month and therefore, both signals—recommendations and short interest—are obtained at approximately the same time during the month. Second, the delay ensures that investors are given ample time to process and impound in price whatever new information is contained in both signals. Thus, we purposely exclude from our tests any drift in stock prices that occurs due to the public disclosure of the signal. In this manner, our methodology differs markedly from the daily rebalancing requirements employed in papers such as Barber et al. (2001) and Barber et al. (2010).

6 Performing these analyses using quarterly data is intuitive given that the majority of the predictor variables (seven of the 11) change on a quarterly basis as financial information is disclosed. We find that all inferences are the same when we perform the analyses using monthly data.

7 Ljungqvist et al. (2009) provide evidence that the I/B/E/S recommendations database contains systematic errors in the pre-2007 files that is likely to overstate the investment value of analysts' recommendations. Our study is among the first to re-examine the investment value of analysts' recommendations using the cleaned 2007 database.

8 Data from the online vendor, shortsqueeze.com, provide less than 1 percent of the total observations. The data cover a period in which we were unable to obtain short interest directly from the NASDAQ. We compared shortsqueeze.com data to that from a six-month period for which we already had short interest data from NASDAQ. The only differences were due to shortsqueeze.com rounding their data to the nearest hundredth place.

9 In the "Robustness Tests" section, we examine the sensitivity of our results to deflating by lagged trading volume.

Similar to JKKL, we select a set of 11 predictor variables for our analysis and winsorize each of the predictor variables at the 2.5 and 97.5 percentiles to control for outliers. We group the predictor variables into one of four classifications based on the nature of the variable (see the Appendix for details on the calculation of each variable). The first group, labeled *Accounting*, consists of earnings surprise (*SUE*), total accruals (*TACCR*), and capital expenditures (*CAPEX*). The *Valuation* group consists of the market-value-of-equity (*MVE*), earnings-to-price ratio (*EP*), book-to-market ratio (*BTM*), and the average daily stock turnover (*TURN*). The *Growth* group consists of realized sales growth (*SG*) and forecasted long-term growth (*LTG*). The fourth group, *Momentum*, consists of earnings forecast revision (*FREV*) and price momentum (*MOM*). These variables have been shown in prior research to be associated with future returns (see the Appendix for specific citations). Thus, we expect that sophisticated capital market participants, such as analysts and short sellers, would use information embedded in these variables when establishing their positions.

Finally, we obtain monthly returns data from CRSP to compute future six-month buy-and-hold abnormal returns (*ARET6*) using characteristic portfolio-matching.[10] Specifically, we define abnormal returns as the raw buy-and-hold return adjusted for the portfolio return from 125 benchmark portfolios formed based on size, book-to-market, and momentum ($5 \times 5 \times 5$), as described in Daniel et al. (1997). We use a holding period of six months for consistency with JKKL, but test the sensitivity of our results to alternative holding periods as a supplemental analysis.

Our final sample, resulting from the intersection of Compustat, CRSP, I/B/E/S, and our short interest database, consists of 80,674 firm-quarter observations over the 52 calendar quarters from 1994 to 2006. For our main analyses, we rank firms into quintile portfolios based on analyst recommendations (both levels and changes) and short interest in each calendar quarter *t*. Thus, we rebalance the portfolios quarterly. For recommendation changes, we ensure that firms without a recommendation revision are included in the middle quintile.

Table 1, Panel A presents descriptive statistics for the variables used in our analyses. The mean (median) value for *Rec* of 3.76 (3.79) indicates that the average analyst recommendation is only moderately less than a "buy" (which would be coded 4). A narrow interquartile range of 0.34 (−0.20 to 0.14) for the consensus recommendation change, *ChgRec*, shows that analyst recommendations are generally sticky. Nevertheless, the minimum and maximum values for *ChgRec* indicate that analysts occasionally downgrade a stock all the way from strong buy to strong sell, and *vice versa*. The mean short interest ratio, *SIratio*, is 3.2 percent, which is considerably larger than the median of 1.8 (due to some large values, as indicated by the maximum of 23.5 percent). The mean six-month abnormal return (*Aret6*) is 1.0 percent, but the median is only −1.9 percent.

With respect to the 11 predictor variables, we find that earnings surprise (*SUE*) has a mean of zero but a slightly positive median of 0.002, consistent with most firms reporting earnings that meet or beat the current analyst forecast. Total accruals (*TACCR*) are negative, on average, due to deducting depreciation. Capital expenditures (*CAPEX*) average approximately 6.3 percent of assets. We find that firm size (*MVE*) is highly skewed, with a mean of $3,578 compared to a median of $775 (in millions). The average earnings-to-price ratio (*EP*) is only 2.0 percent due to some negative values (median 4.3 percent). The book-to-market ratio (*BTM*) has a mean of 0.50 (median 0.42), consistent with prior research. Approximately 0.58 percent of a firm's shares turn over on any given day (*TURN*). Realized sales growth (*SG*) averages 17 percent, and analysts' long-term earnings growth forecasts (*LTG*) average 17.55 percent. Analysts' forecast revisions (*FREV*) have

[10] If a firm delists during the return accumulation period, we compound the delisting return with the buy-and-hold return and assume the liquidating proceeds are reinvested in a portfolio that earns a normal return for the remainder of the period.

TABLE 1

Descriptive Statistics

Panel A: Summary Statistics for Dependent and Explanatory Variables

Variable	Mean	Std. Dev.	Min	Q1	Median	Q3	Max
Rec	3.76	0.62	1.00	3.33	3.79	4.17	5.00
ChgRec	−0.03	0.42	−4.00	−0.20	0.00	0.14	4.00
SIratio	3.2%	3.8%	0.0%	0.6%	1.8%	4.2%	23.5%
Aret6	0.010	0.371	−1.000	−0.177	−0.019	0.144	12.525
SUE	0.000	0.029	−0.237	−0.004	0.002	0.006	0.312
TACCR	−0.012	0.035	−0.151	−0.028	−0.011	0.005	0.092
CAPEX	0.063	0.061	0.000	0.022	0.045	0.082	0.341
MVE	3,578	8,421	26	255	775	2,512	65,417
EP	0.020	0.118	−1.553	0.016	0.043	0.065	0.214
BTM	0.500	0.343	0.045	0.262	0.424	0.648	2.647
TURN	0.582	0.273	0.000	0.364	0.614	0.819	1.000
SG	1.17	0.26	0.47	1.02	1.11	1.25	2.35
LTG	17.55	8.02	4.25	12.00	15.44	21.07	55.00
FREV	0.000	0.029	−0.163	−0.005	0.002	0.010	0.219
MOM	0.084	0.421	−0.995	−0.125	0.047	0.228	27.827

Panel B: Analyst Recommendations and Short Interest Values by Calendar Year

Year	n	Mean *Rec*	Mean *SIratio*	Returns for S&P 500
1994	4,319	3.74	1.6%	1.3%
1995	4,586	3.76	1.7%	37.6%
1996	5,046	3.83	1.8%	23.0%
1997	5,554	3.89	2.2%	33.4%
1998	6,021	3.90	2.3%	28.6%
1999	6,514	3.93	1.9%	21.0%
2000	6,099	4.00	2.0%	−9.1%
2001	5,772	3.87	2.9%	−11.9%
2002	6,559	3.73	3.6%	−22.1%
2003	7,403	3.52	4.0%	28.7%

(continued on next page)

Panel B: Analyst Recommendations and Short Interest Values by Calendar Year

Year	n	Mean *Rec*	Mean *SIratio*	Returns for S&P 500
2004	7,686	3.65	4.0%	10.9%
2005	7,663	3.61	4.6%	4.9%
2006	7,452	3.63	5.7%	15.8%
Full Sample	80,674	3.76	3.2%	12.5%

The sample consists of 80,674 firm-quarter observations during the period 1994–2006. See the Appendix for a more detailed description of how each variable is calculated.

Variable Definitions:

Rec = consensus analyst recommendation in the last month of the calendar quarter, where 5 = strong buy, 4 = buy, 3 = hold, 2 = sell, and 1 = strong sell. We calculate *Rec* as the mean of all outstanding recommendations issued a maximum of 12 months prior to month-end. In calculating means, we use only the most recent individual analyst recommendation issued on or before the I/B/E/S consensus recommendation date;
ChgRec = change in the consensus analyst recommendation from the previous quarter;
SIratio = number of shares sold short as reported for the last month of the calendar quarter divided by the number of shares outstanding as of the same date;
Aret6 = six-month abnormal return (adjusted for firm size, book-to-market ratio, and past returns) beginning the first day of quarter t+1;
SUE = seasonally adjusted earnings change scaled by price for fiscal quarter q;
TACCR = total accruals scaled by average assets measured at the end of fiscal quarter q;
CAPEX = rolling sum of the preceding four quarters of capital expenditures ending at fiscal quarter q divided by total assets;
MVE = market value of equity at the end of fiscal quarter q;
EP = ratio of the rolling sum of earnings over the preceding four quarters to price at the end of fiscal quarter q;
BTM = ratio of book value of equity to market value of equity as of the end of fiscal quarter q;
TURN = average daily volume per share over the preceding six months;
SG = rolling sum of sales growth over the preceding four fiscal quarters;
LTG = consensus long-term earnings growth forecast at the end of calendar quarter t;
FREV = rolling sum of the preceding six-month earnings forecast revisions scaled by price; and
MOM = price momentum, measured as the six-month raw return ending one month prior to the end of the fiscal quarter q.

a mean of zero but a slightly positive median of 0.002. Price momentum (*MOM*) averages 8.4 percent for the preceding six months (median 4.7 percent).

Table 1, Panel B reports mean analyst recommendations, short interest, and market returns (using the S&P 500 portfolio) for each year from 1994 to 2006. From 1994 to 2000, we observe a monotonic increase in the average analyst recommendation, which peaks at 4 in 2000. The average recommendation then declines in years 2001 through 2003, and remains at a lower level through the end of our test period. This shift corresponds with criticism of analysts that led to the Global Research Analysts Settlement, NASD 2711, and NYSE Rule 472. One line of criticism focused on analysts' conflicts of interest, including their incentive to maintain a positive relation with corporate managers in order to generate investment banking business and to obtain earnings guidance.

Table 1, Panel B also reports another noteworthy change over our test period. The mean level of short interest is around 2 percent from 1994 to 2000. The level then increases appreciably over the next six years, reaching 5.7 percent in the final year of our sample period. This shift, which corresponds with a dramatic increase in the number of hedge funds, increases the importance of research that furthers an understanding of the role of short selling in the price formation process. Note that shifts over time have a minimal effect on our results because we rank firms into quintiles based on their relative values at a given point in time.

III. PREDICTIVE INFORMATION USED BY ANALYSTS AND SHORT SELLERS

In this section, we first examine whether analyst recommendations and short interest incorporate fundamental and other information in the manner shown by prior research to be predictive of future returns. We then investigate whether analyst recommendations and short interest provide information that is incremental to that information.

Univariate Evidence

Table 2 presents mean values for each of the 11 predictive variables by quintile for recommendation levels, recommendation changes, and short interest levels. The quintiles correspond to portfolios that we later use in trading analyses. In Panel A, as we move down each column from the worst to the best recommendations, we find a monotonic (or near monotonic) increase for eight of the 11 variables. The increase for *SUE*, *EP*, *FREV*, and *MOM* is consistent with analyst recommendations properly incorporating the relation of these measures with future returns. In contrast, the increase for *TACCR*, *CAPEX*, *SG*, and *LTG* indicate that analysts misuse this information, which could cause more favorable recommendations to portend lower investment returns. The overall pattern of information use indicates that analysts tend to issue more favorable recommendations for glamour stocks, even though prior studies show that these stocks earn lower subsequent returns (Lakonishok et al. 1994; La Porta 1996; Sloan 1996; Beneish et al. 2001). Examining changes in recommendations, Panel B shows a clear pattern for only three variables; but in each case, the change is consistent with the relation of the information with future returns established in prior research. Specifically, as we move down the columns from downgrades to upgrades, we observe a monotonic increase for earnings forecast revisions (*FREV*) and stock price momentum (*MOM*), and a monotonic decrease for long-term growth (*LTG*). While prior research has generally found that recommendation revisions are better predictors of future returns than are recommendation levels, this analysis indicates that recommendation revisions fail to incorporate eight of the 11 items of predictive information. These results for recommendation levels and changes are similar to the results documented in JKKL.

Panel C of Table 2 provides the corresponding analysis for short interest quintiles. The book-to-market variable (*BTM*) decreases monotonically as the level of short interest increases, so short sellers tend to take greater positions in firms with a higher market value relative to their book

TABLE 2

Mean Values by Quintile for 11 Variables Associated with Future Returns

Panel A: Recommendation Levels

	QRec	n	*Rec*	*SUE*	*TACCR*	*CAPEX*	*MVE*	*EP*	*BTM*	*TURN*	*SG*	*LTG*	*FREV*	*MOM*
Worst	0.00	16,216	2.92	−0.005	−0.015	0.057	1,923	−0.010	0.659	0.523	1.082	15.16	−0.008	−0.014
	0.25	15,857	3.45	−0.002	−0.013	0.063	4,185	0.020	0.512	0.595	1.134	16.30	−0.002	0.041
	0.50	17,119	3.80	0.001	−0.012	0.065	4,716	0.029	0.468	0.586	1.164	17.42	0.001	0.088
	0.75	15,704	4.07	0.003	−0.010	0.065	4,763	0.031	0.417	0.621	1.212	18.75	0.004	0.139
Best	1.00	15,778	4.59	0.003	−0.009	0.065	2,256	0.031	0.443	0.584	1.237	20.20	0.006	0.171

Panel B: Recommendation Changes

	QChgRec	n	*ChgRec*	*SUE*	*TACCR*	*CAPEX*	*MVE*	*EP*	*BTM*	*TURN*	*SG*	*LTG*	*FREV*	*MOM*
Down	0.00	16,113	−0.18	0.000	−0.011	0.064	2,711	0.025	0.503	0.597	1.184	17.80	−0.004	0.026
	0.25	13,869	−0.05	0.000	−0.013	0.068	5,633	0.021	0.445	0.653	1.188	17.75	0.000	0.062
No	0.50	22,799	0.00	−0.001	−0.012	0.059	1,903	0.014	0.574	0.497	1.142	17.54	0.000	0.072
	0.75	11,690	0.04	0.001	−0.012	0.066	6,328	0.025	0.421	0.651	1.182	17.46	0.003	0.130
Up	1.00	16,070	0.15	0.001	−0.012	0.063	3,059	0.021	0.499	0.574	1.148	17.20	0.003	0.147

Panel C: Short Interest Levels

	QSIratio	n	*SIratio*	*SUE*	*TACCR*	*CAPEX*	*MVE*	*EP*	*BTM*	*TURN*	*SG*	*LTG*	*FREV*	*MOM*
Low	0.00	16,118	0.3%	0.000	−0.012	0.059	3,643	0.022	0.622	0.346	1.120	16.10	0.001	0.065
	0.25	16,144	1.1%	0.001	−0.012	0.059	5,287	0.028	0.501	0.492	1.129	15.77	0.001	0.083
	0.50	16,145	2.0%	0.000	−0.012	0.061	4,156	0.025	0.474	0.588	1.142	16.63	0.001	0.096
	0.75	16,144	3.5%	0.000	−0.012	0.065	2,972	0.018	0.459	0.686	1.177	18.28	0.000	0.094
High	1.00	16,123	8.9%	−0.001	−0.011	0.072	1,832	0.009	0.446	0.796	1.259	20.97	−0.002	0.083

The sample consists of 80,674 firm-quarter observations during the period 1994–2006. See Table 1 for descriptions of each variable, and the Appendix for detailed explanations of how each variable is calculated.

value. Three variables increase monotonically as the level of short interest increases: capital expenditures (*CAPEX*), stock turnover (*TURN*), and sales growth (*SG*). In each case, the pattern of information use is consistent with their relation to future returns documented in prior research. Comparing the high and low short interest quintiles for the other variables shows that short positions are generally consistent with how the variables map into future returns, but the relation is not monotonic.

Multivariate Evidence

We next use ordered logistical regression analysis to provide a multivariate test of the relation between analyst and short seller investment signals and the 11 predictor variables. In all regression analyses, we assess statistical significance using test statistics based on standard errors that are adjusted for two-way clustering of residuals by firm and calendar month (Petersen 2009; Gow et al. 2010). Table 3, Panel A reports results using analyst recommendations and recommendation revision quintiles as the dependent variable, with quintiles coded from 1 to 5.[11]

For recommendation levels, we find that analysts correctly incorporate the implications for future returns of only one of the *Accounting* variables: unexpected earnings (*SUE*). Analysts favorably recommend firms with high total accruals (*TACCR*), and do not consider capital expenditures (*CAPEX*), despite evidence that increases in those accounting measures are associated with lower future returns (Sloan 1996). Examining the *Valuation* measures, analysts correctly favor smaller firms (*LNMVE*) and those with a higher earnings-to-price ratio (*EP*). However, they also favor firms with a low book-to-market ratio (*BTM*) and high growth (*SG*, *LTG*), despite evidence that stock prices of such firms underperform the market. Examining the *Momentum* variables, analysts correctly favor firms with high earnings momentum (*FREV*) and stock price momentum (*MOM*).

The results for revisions in analysts' recommendations are reported on the right side of Table 3, Panel A. Examining the *Accounting* variables, we find that all three variables (*SUE*, *TACCR*, and *CAPEX*) are statistically significant, but with the unexpected sign. For the *Valuation* and *Growth* variables, the evidence is mixed: *EP* and *TURN* are statistically significant in the expected direction, but *BTM* is statistically significant in the unexpected direction. The coefficient on *LTG* is also significant in the unexpected direction. For the *Momentum* variables, we find that both *MOM* and *FREV* are statistically significant in the expected direction. Finally, we find that recommendation changes are negatively associated with past recommendation levels. This result makes intuitive sense because the highest (lowest) recommendations can only be revised down (up).

Considering the types of information used by analysts in both their recommendations and recommendation revisions, analysts' correctly favor stocks with positive price momentum (*MOM*), positive earnings momentum (*FREV*), and high earnings-to-price (*EP*). They incorrectly favor stocks with high forecasted growth (*LTG*), high accruals (*TACCR*), and low book-to-market value (*BTM*). Thus, financial analysts view higher past and future growth and higher accruals as positive features in recommending stocks, despite research that shows the opposite relation (Lakonishok et al. 1994; La Porta 1996; Sloan 1996). In addition, analysts also tend to issue more favorable recommendations for firms with low book-to-market ratios, even though prior research shows a positive association with subsequent returns (Fama and French 1992). This evidence indicates that

[11] Note that quintiles are of approximate equal size (after adjusting for ties and including all recommendation revisions of zero in the middle quintile). Due to the low frequency of strong sell and sell recommendations issued by analysts, the most unfavorable recommendation quintile contains some "hold" recommendations.

TABLE 3

Use of Predictive Information by Analysts and Short Sellers

Panel A: Explaining Recommendation Levels and Changes (Using Ordered Logistic Regression)

		Recommendation Levels		**Recommendation Changes**	
Variable	**Predict**	**Coefficient**	**Chi-Square**	**Coefficient**	**Chi-Square**
Accounting					
SUE	Pos	1.837	15.70***	−0.743	13.81***
TACCR	Neg	0.994	21.16***	0.424	4.52**
CAPEX	Neg	0.006	0.00	0.437	8.42***
Valuation					
LnMVE	Neg	−0.034	8.78***	0.008	1.34
EP	Pos	2.278	73.73***	0.564	20.31***
BTM	Pos	−0.714	119.05***	−0.170	31.79***
TURN	Neg	−0.032	0.17	−0.319	58.70***
Growth					
SG	Neg	0.781	213.28***	−0.015	0.15
LTG	Neg	0.041	202.87***	0.015	61.73***
Momentum					
FREV	Pos	8.254	115.68***	5.170	76.91***
MOM	Pos	0.473	33.60***	0.499	71.64***
LAG_QRec	Neg			−0.926	1052.85***
Pseudo R^2		0.138		0.098	

(continued on next page)

Panel B: Explaining Short Interest (Using Ordered Logistic Regression)

Variable	Predict	Short Interest Coefficient	Chi-Square
Accounting			
SUE	Neg	−1.777	21.93***
TACCR	Pos	1.527	21.63***
CAPEX	Pos	1.974	32.66***
Valuation			
LnMVE	Pos	0.080	2.83*
EP	Neg	−0.698	18.04***
BTM	Neg	−0.245	14.46***
TURN	Pos	4.198	484.78***
Growth			
SG	Pos	0.410	45.46***
LTG	Pos	0.014	23.66***
Momentum			
FREV	Neg	−2.138	15.27***
MOM	Neg	−0.096	5.25**
Pseudo R^2		0.332	

*, **, *** Indicate statistical significance at the $\alpha = 0.10$, 0.05, and 0.01 levels, respectively, using a two-tailed test.

n = 80,674 firm-quarters.

This table reports estimation results when analysts' recommendation and short interest quintile assignments are regressed (using ordered Logit) on 11 variables shown to be predictive of future returns. We do not report the intercepts for parsimony. See Table 1 for descriptions of each variable, and the Appendix for detailed explanations of how each variable is calculated. The "Predict" column reports the predicted relation between the explanatory variable and future returns as indicated in prior research. We report test statistics based on standard errors that are adjusted for two-way clustering of residuals by firm and calendar month.

sell-side analysts tend to favorably recommend "glamour stocks." JKKL reached the same conclusion based on an analysis of an earlier time period.

Table 3, Panel B reports results from a model using short interest quintiles as the dependent variable. We find that all 11 variables are statistically significant with coefficient signs in the expected direction.[12] Additionally, we note that the explanatory power of this model (Panel B, pseudo $R^2 = 33.2$ percent) is more than double that for the model using recommendation levels (Panel A, pseudo R^2 of 13.8 percent) and more than triple that for the model using recommendation revisions (Panel A, pseudo R^2 of 9.8 percent).[13] Thus, consistent with our univariate analysis, we find that short interest is explained better by the predictive information in *Accounting*, *Valuation*, and *Growth* variables than is analyst recommendation levels or changes. Our evidence is consistent with other studies that examine the association between short interest and indicators of future returns (Dechow et al. 2001; Cao et al. 2007; Seybert and Wang 2009). Our study extends this research by employing several predictive variables simultaneously in the same regression model, and by comparing results for short sellers to those for analysts.

Incremental Information About Future Returns

In this section, we investigate whether recommendations, recommendation changes, and short interest contain incremental information about future returns, beyond the information in the 11 predictive variables. Using the methodology in JKKL, we convert the continuous predictor variables into binary signals based on a median split. For all variables where the expected relation with future returns is positive (negative), the binary variable is coded 1 when its value is greater (less) than its median for a given quarter, and 0 otherwise. Thus, we expect a positive coefficient on all predictor variables.

Table 4 presents the results from our analysis. The model on the left includes both recommendation level quintiles (*QRec*) and short interest quintiles (*QSIratio*). Each of these quintiles is scaled to range from 0 to 1 to facilitate interpretation of coefficients. The coefficient on *QRec* indicates that analysts' recommendations are incrementally informative about future abnormal returns, but the coefficient is negative. This result suggests that buy-and-hold investors would do better to trade against the consensus analyst recommendation. In contrast, the coefficient on *QSIratio* is significantly negative, indicating that short interest provides incremental information for predicting future returns even after controlling for the information contained in the 11 predictor variables. The negative sign indicates that, as would be expected, a higher (lower) level of short interest is associated with lower (higher) future abnormal returns. The coefficient magnitude for *QSIratio* of -0.029 can be interpreted as the six-month return earned on an investment portfolio that is formed optimally to exploit the information in short interest that is orthogonal to the information in analysts' recommendations and the predictive variables. Thus, while short sellers use the information contained in the investment signals (as indicated by the significant associations reported in Table 3), the results in Table 4 show that short sellers also develop information to predict future returns that goes beyond what is contained in those variables. Examining the predictor variables, we find significant coefficients in the expected direction for *TACCR*, *SG*, and *FREV*.[14]

---

[12] Note that the explanatory variables have the opposite predicted sign in the short interest model (compared to the recommendation models).

[13] Since the dependent variables differ across models, it is not possible to test for differences in explanatory power. However, given that we have standardized the dependent variables by ranking them into quintiles, their variation is similar. Specifically, the standard deviations of the quintile ranking of analyst levels, analyst changes, and short interest are 1.41, 1.38, and 1.41, respectively. Thus, we believe a comparison of pseudo R^2s is informative.

[14] The coefficient for *TURN* is marginally significant with the wrong sign. This appears to be driven by its high correlation with short interest (greater than 50 percent). When *QSIratio* is excluded from the model, the coefficient for *TURN* is not

TABLE 4

Incremental Information about Future Returns Provided by Recommendations, Recommendation Revisions, and Short Interest

Variable	Recommendations and Short Interest		Recommendation Changes and Short Interest	
	Coefficient	t-stat	Coefficient	t-stat
Intercept	0.009	0.70	−0.003	−0.31
QRec	−0.019	−2.55**		
QChgRec			0.003	0.59
QSIratio	−0.029	−3.91***	−0.027	−3.77***
DSUE	0.004	0.79	0.003	0.63
DTACCR	0.025	7.10***	0.025	7.23***
DCAPEX	0.004	0.75	0.004	0.72
DLnMVE	0.003	0.72	0.003	0.68
DEP	−0.003	−0.45	−0.004	−0.54
DBTM	0.005	0.93	0.007	1.19
DTURN	−0.012	−1.85*	−0.012	−1.80*
DSG	0.015	2.43**	0.016	2.66***
DLTG	−0.009	−1.08	−0.007	−0.85
DFREV	0.011	2.05**	0.009	1.75*
DMOM	0.009	1.13	0.007	0.94
Adj-R^2	0.003		0.003	

*, **, *** Indicate statistical significance at the α = 0.10, 0.05, and 0.01 levels, respectively, using a two-tailed test.

n = 80,674 firm-quarters.

This table reports estimation results when future six-month abnormal returns (calculated as in Daniel et al. [1997]) are regressed on analysts' recommendations and short interest data along with 11 variables that prior research shows to be predictive of future returns. *QRec* is the quintile assignment based on recommendation levels. *QChgRec* is the quintile assignment based on recommendation revisions. *QSIratio* is the quintile assignment based on short interest. *QRec, QChgRec,* and *QSIratio* are scaled to range between 0 and 1 (0.00, 0.25, 0.50, 0.75, 1.00) to facilitate the interpretation of the coefficients. Using the methodology in Jegadeesh et al. (2004), we convert the continuous predictor variables into binary signals based on a median split. For all variables where the expected relation with future returns is positive (negative), the binary variable is coded 1 when its value is greater (less) than its median for a given quarter, and 0 otherwise. See Table 1 for descriptions of each variable, and the Appendix for detailed explanations of how each variable is calculated. We report test statistics based on standard errors that are adjusted for two-way clustering of residuals by firm and calendar month.

The model on the right side of Table 4 provides a similar analysis for recommendation revision quintiles (*QChgRec*) and short interest (*QSIratio*) quintiles. The coefficient on *QChgRec* is positive as expected, but is not significantly different from zero, suggesting that recommendation revisions do not provide information about future returns that is incremental to short interest and to the other publicly available investment signals. The coefficient on *QSIratio* is again significantly negative. The coefficients and significance levels for the predictor variables are similar to those reported in the recommendation level regressions.[15]

statistically significant (t-stat = −0.52). In contrast, *QSIratio* remains highly significant when *TURN* is excluded from the model.

[15] We also estimated a regression equation that includes *QRec*, *QChgRec*, and *QSIratio* together with the other 11 predictive variables. Results from this regression are qualitatively equivalent to what is reported in Table 4, except that

We draw the following general conclusions from the results presented in Table 3 and Table 4. First, analysts' recommendations do a poor job of incorporating information about future returns provided by the predictor variables, and buy-and-hold investors would actually do better by trading against analyst recommendations. Second, analyst recommendation revisions also do a poor job of using predictive information, and they do not contribute information beyond what is contained in the predictor variables and short interest. Third, short sellers correctly incorporate publicly available information that is predictive of future returns and, furthermore, short sellers are able to generate information that is orthogonal to the set of predictive variables we use in our analysis. In the next section, we explore the success of trading strategies that are designed to exploit the above results.

IV. INVESTMENT PERFORMANCE BASED ON ANALYST RECOMMENDATIONS AND SHORT INTEREST

The results reported in Section III suggest that investors might improve their returns by using short interest as a supplementary investment signal. Indeed, many of the associations between analyst recommendations and variables that are predictive of future returns suggest that analysts might actually impede the price discovery process. Pownall and Simko (2005) provide evidence that short positions can help to fill an information intermediary gap for companies with low analyst following. Our results suggest that short interest can be used as an information signal for investors, regardless of the analyst following. The benefit of short interest is likely to be greatest when this signal contradicts the signal from analysts' recommendations. Specifically, the results in Table 4 suggest that investors would profit from a buy-and-hold strategy that (1) trades against analyst recommendation levels and (2) trades with short interest. This conjecture is based on the signs of the statistically significant coefficients for analyst recommendations and short interest in Table 4.[16]

The trading strategies we consider are implementable and follow the portfolio construction methodology outlined in Jegadeesh and Titman (1993).[17] Under this methodology, the strategies hold a series of sub-portfolios that formed in the current month and in each of the previous five months (six-month holding period).[18] Thus, we simulate a portfolio where a 1/6 fraction of the stocks are reassigned to portfolios each month. We rebalance the portfolios monthly to maintain equal weights on each security and calculate the mean abnormal return for each portfolio. This results in a time-series of monthly portfolio returns that is free of overlapping return accumulation periods. As discussed in further detail below, we also calculate hedge portfolio returns that go long and short in particular portfolios.

Since the consensus analyst recommendation and the short interest ratio change each month, each portfolio is based on information from analysts and/or short sellers from the current month and from each of the past five months. In our initial tests, we examine the signals from analysts or short sellers separately. We then combine recommendations and short interest to develop a trading strategy that exploits the information contained in both signals. When combining the signals, we intersect the analyst recommendation and short interest quintiles to produce 25 portfolios (5×5).

the coefficient on *QChgRec* becomes marginally significantly positive.

16 In this section, we do not report trading strategies that use recommendation revisions because results from these tests are generally consistent with our earlier analyses that show they contain no incremental information about future returns beyond short interest.

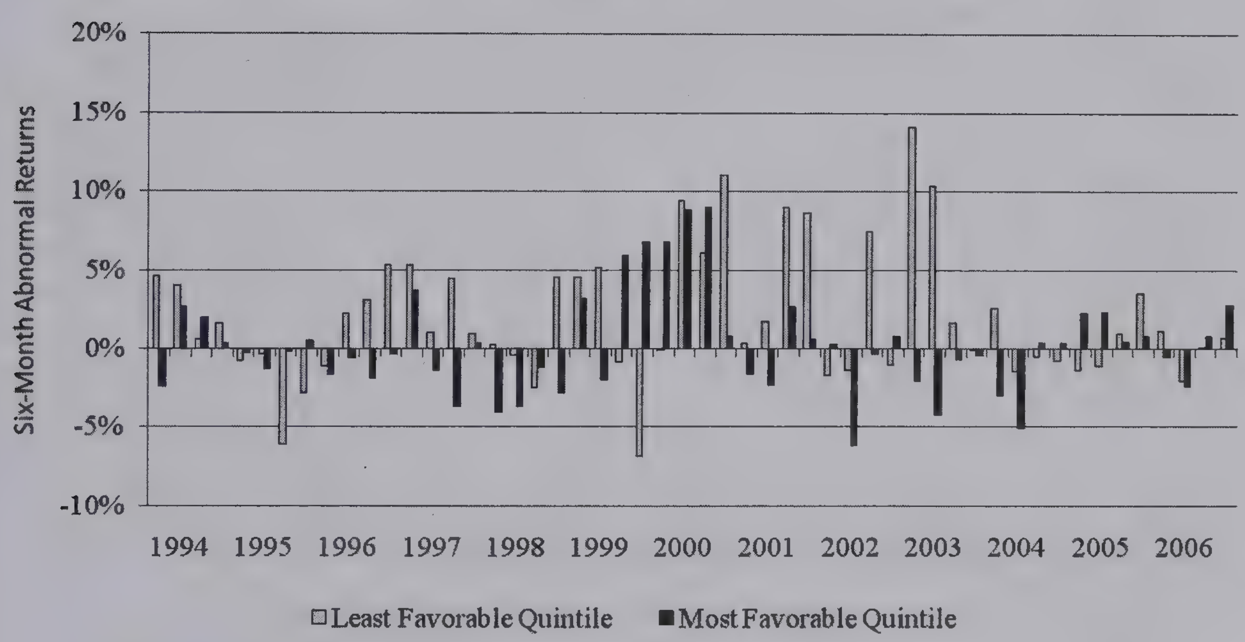
17 The Jegadeesh and Titman (1993) methodology avoids statistical problems associated with serial correlation induced by overlapping return accumulation periods. We thank an anonymous reviewer for suggesting this approach.

18 In a later section, we discuss the sensitivity of the results to a shorter return window of one month and the use of a four-factor model, which controls for the Fama-French risk factors and momentum.

We form our portfolios using all firm-month observations with available data for analyst recommendations, short interest levels, and stock returns. Thus, we drop the requirement that data are available for the 11 predictor variables, and we include every month (not just the last month of each quarter, as in our first set of analyses). These changes increase the sample size to 564,101 firm-month observations and, as a result, increase the extent to which we can generalize results.[19] We report results for the full time period, 1994 to 2006, and for three sub-periods because documenting the stability of abnormal returns to a particular trading strategy over time is critical to the evaluation of the strategy. The first sub-period consists of the five-year period from 1994 to 1998, which overlaps the 1985 to 1998 time period investigated by JKKL, and corresponds to a strong bull market. The second sub-period consists of the subsequent five years, from 1999 to 2003, which includes the precipitous decline in the NASDAQ index, the adoption of Regulation Fair Disclosure, and the Global Settlement agreement reached between the SEC, NASD, NYSE, and ten of the largest investment firms. The final sub-period consists of the most recent years in our sample, 2004 to 2006.

In Figure 1, we display six-month abnormal quarterly returns over the full test period for

FIGURE 1
Comparison of Six-Month Abnormal Returns to the Most and Least Favorable Recommendations



Consensus recommendations as of the last month of calendar quarter t are sorted into quintiles, with the highest quintile designated the most favorable portfolio and the lowest quintile the least favorable portfolio. Portfolios are reformed each calendar quarter and we report abnormal buy-and-hold returns for the six months beginning the first day of quarter $t+1$.

[19] In a robustness test (untabulated), we find that inferences from our trading strategies are unchanged when we restrict the sample to include firms *with* available data for the 11 predictor variables.

stocks in the least favorable recommendation quintile and for those in the most favorable recommendation quintile. Our first observation is that the variability of the absolute value of the abnormal returns differs considerably among the three sub-periods. The variability is much higher in the 1999–2003 sub-period than in the preceding 1994–1998 sub-period; return variability then drops precipitously and remains relatively low throughout the 2004–2006 sub-period. Our second observation is that analyst recommendations are not very reliable as a predictor of future returns. In fact, stocks in the least favorable recommendation quintile earn greater returns than those in the most favorable quintile in 35 of the 72 quarters. So, flipping a coin appears to be just as predictive.

Returns for Portfolios Based on Analyst Recommendations or Short Interest

Table 5, Panel A reports average monthly abnormal returns based on analyst recommendations. The portfolios cover the 52 calendar quarters from 1994 to 2006. If analyst recommendations provide value to investors, then returns should be higher for more highly recommended stocks. That is, returns for stocks in the quintile with the highest average recommendation should exceed returns for lower recommendation quintiles. Consistent with the results previously presented in Table 4 (which indicate that analysts improperly incorporate value-relevant information into their forecasts), we find the opposite relation. We find that the mean abnormal returns generally *decrease* as the recommendation level increases. When we calculate returns from investing long in firms classified in the most favorable recommendation quintile and taking an offsetting short position in firms in the least favorable recommendation quintile, we obtain a *modestly negative* return of 26 monthly basis points ($t = -2.03$; $p < 0.01$).

Table 6, Panel A summarizes returns by recommendation quintile for each of the three sub-periods and provides statistical tests. The investment results vary considerably among the three sub-periods. However, the most favorable recommendation quintile does not perform better than fourth among the quintiles (see ranks in the right-most column). Two statistically significant portfolios fall in the 1999–2003 period, when returns are positive for the lowest recommendation quintiles. We also find that the hedge portfolio trading strategy (buying firms in the highest recommendation quintile and selling short those in the lowest recommendation quintile) yields a statistically significant return only in the 1999–2003 sub-period, and that return is a *negative* 48 monthly basis points. Overall, following the consensus analyst recommendation does not generate positive abnormal returns for investors and, in some periods, can actually generate significantly negative returns. While perhaps counterintuitive, this result is consistent with Barniv et al. (2009), who find that analysts' recommendations relate negatively to residual income valuation models, and with Seybert and Wang (2009), who find that firms with more optimistic recommendations earn lower future returns in periods of high investor sentiment.[20]

Table 5, Panel B reports average monthly abnormal returns from portfolios formed using short interest levels. We find that the average abnormal return across the quintiles declines monotonically from lowest to highest short interest. A trading strategy that invests long (short) in firms that are in the lowest (highest) short interest quintile provides an average monthly abnormal return of 56 basis points, which is both economically and statistically significant ($t = 2.46$; $p < 0.01$). Interestingly, 29 basis points of that return are earned because short sellers avoid sizable positions in stocks that yield positive future returns. Thus, not only are short sellers able to identify stocks that are likely to fall in price, but they successfully avoid stocks that are likely to realize price

[20] Relative to JKKL, we find a more negative relation between analyst recommendations (both levels and changes) and future returns. Our sub-period analysis suggests that an important factor driving this result is the different time periods examined across the two studies.

TABLE 5

Abnormal Returns to Portfolios Based on Analyst Recommendations and Short Interest 1994–2006

Panel A: Average Monthly Abnormal Returns Based on Recommendation Levels

	Performance 1994 to 2006		
Quintile Portfolios	**Mean Return**	**t-stat**	**Quintile Rank by Return**
Most Favorable Recommendations	−0.10%	−1.37	5 (Worst)
	0.01%	0.13	4
	0.11%	2.32**	3
	0.19%	3.64***	1 (Best)
Least Favorable Recommendations	0.16%	2.15**	2
Most Favorable – Least Favorable	−0.26%	−2.03**	

Panel B: Average Monthly Abnormal Returns Based on Short Interest

	Performance 1994 to 2006		
Quintile Portfolios	**Mean Return**	**t-stat**	**Quintile Rank by Return**
Lowest Short Interest	0.29%	3.19***	1
	0.20%	3.64***	2
	0.12%	2.70***	3
	0.04%	0.63	4
Highest Short Interest	−0.27%	−1.79*	5
Lowest SI – Highest SI	0.56%	2.46**	

*, **, *** Indicate statistical significance at the $\alpha = 0.10$, 0.05, and 0.01 levels, respectively, using a two-tailed test.

n = 564,101 firm-months.

We report mean monthly abnormal returns adjusted for firm size, book-to-market ratio, and past returns (Daniel et al. 1997) for portfolios based on analyst recommendations or short interest. Using the methodology in Jegadeesh and Titman (1993), we construct portfolios monthly based on the quintile rank of analyst recommendations or short interest selected in the current month and in each of the past five months (six-month holding period).

TABLE 6

Abnormal Returns to Portfolios Based on Analyst Recommendations or Short Interest for Sub-Periods

Panel A: Average Monthly Abnormal Returns Based on Recommendation Levels

	Performance 1994 to 1998			Performance 1999 to 2003			Performance 2004 to 2006		
Quintile Portfolios	**Mean Return**	**t-stat**	**Rank by Return**	**Mean Return**	**t-stat**	**Rank by Return**	**Mean Return**	**t-stat**	**Rank by Return**
Most Favorable Recommendations	−0.16%	−1.53	5	−0.17%	−1.22	5	0.02%	0.26	4
	−0.07%	−0.58	4	0.03%	0.23	4	0.05%	0.75	3
	0.11%	1.68*	2	0.15%	1.53	3	0.06%	1.00	1
	0.10%	1.52	3	0.40%	3.93***	1	0.06%	0.80	2
Least Favorable Recommendations	0.13%	1.16	1	0.31%	2.11**	2	−0.06%	−0.63	5
Most Favorable – Least Favorable	−0.29%	−1.58		−0.48%	−1.93*		0.09%	0.55	

Panel B: Average Monthly Abnormal Returns Based on Short Interest

	Performance 1994 to 1998			Performance 1999 to 2003			Performance 2004 to 2006		
Quintile Portfolios	**Mean Return**	**t-stat**	**Rank by Return**	**Mean Return**	**t-stat**	**Rank by Return**	**Mean Return**	**t-stat**	**Rank by Return**
Lowest Short Interest	0.22%	1.72*	1	0.30%	1.69*	1	0.30%	2.74***	1
	0.17%	2.09**	2	0.27%	2.52**	2	0.11%	1.61	2
	0.05%	0.93	3	0.19%	2.17**	3	0.05%	0.79	3
	0.01%	0.10	4	0.18%	1.31	4	−0.08%	−0.97	4
Highest Short Interest	−0.30%	−1.43	5	−0.24%	−0.80	5	−0.24%	−1.59	5
Lowest SI – Highest SI	0.52%	1.63		0.54%	1.20		0.54%	2.43***	

*, **, *** Indicate statistical significance at the α =0.10, 0.05, and 0.01 levels, respectively, using a two-tailed test.

n = 564,101 firm-months.

We report mean monthly abnormal returns adjusted for firm size, book-to-market ratio, and past returns (Daniel et al. 1997) for portfolios based on analyst recommendations or short interest. Using the methodology in Jegadeesh and Titman (1993), we construct portfolios monthly based on the quintile rank of analyst recommendations or short interest selected in the current month and in each of the past five months (six-month holding period).

increases. This finding is consistent with concurrent research by Boehmer et al. (2010), who also find that the positive signals from low short interest can be equal to or greater in absolute magnitude than the negative signal from high short interest.

Table 6, Panel B presents returns from trading on short interest levels for each of the three sub-periods. A monotonic pattern of decreasing returns occurs within each of the three sub-periods as one reads down the table from lowest to highest short interest, with several of the individual quintiles statistically significant. Over the 2004–2006 sub-period, the zero-investment hedge strategy produces a statistically significant abnormal return of 54 monthly basis points. The monthly hedge returns are similar for the other sub-periods but are not statistically significant. Thus, while larger abnormal returns can be earned by trading on the level of short interest rather than on analysts' recommendations, the hedge strategy results are not of sufficient magnitude to be statistically significant in each sub-period. Interestingly, the returns to buying stocks in the lowest short interest quintile are more likely to be statistically significant than are the returns to selling short the stocks in the high short interest quintile. Overall, short interest provides a more reliable signal about which stocks to *buy* than about which stocks should be sold short.

Returns for Portfolios Combining Analyst Recommendations and Short Interest Levels

The evidence presented to this point indicates that none of the signals in isolation produce consistent results across each sub-period. In this section, we investigate whether combining the signals from analysts and short sellers can improve upon this investment performance. We consider trading strategies that are based on concurring and conflicting signals from analysts and short sellers. To conduct our analysis, we independently sort analysts' recommendations and short interest into quintiles and merge these quintile rankings to form 25 different portfolios. Since the sorts are independent, the average number of stocks across portfolios is not equal. Nevertheless, as reported in Table 7, the sample is broadly distributed across the 25 different portfolios, ranging from an average number of firms per portfolio of 492 to 950. This suggests that analysts and short sellers are just as likely to disagree as they are to agree on the future prospects of any one firm.

Table 7 presents abnormal returns for portfolios formed using both analyst recommendations and short interest. Reading down each recommendation column in Table 7, Panel A shows that abnormal returns tend to decline as the level of short interest increases. Notably, the decline is monotonic for the three quintiles with the most favorable recommendations. In addition, reading across each short interest row from left to right shows a general tendency for abnormal returns to decrease as the recommendation becomes more favorable. The combined effect is that the lowest returns occur in the lower right quadrant and the highest in the upper left quadrant.

Table 7, Panel B reports returns for two zero-investment trading strategies. The first strategy is to trade when short sellers and analysts strongly concur about a company's prospects. The specific trade is to buy firms in the portfolio with the most favorable recommendations and lowest short interest, and to sell short firms with the worst recommendations and highest short interest. The upper half of Panel B reports the returns and t-statistics from this strategy. The sample period return is 36 monthly basis points, which is not statistically significant ($t = 1.25$; $p = 0.21$). Examining sub-periods, a positive abnormal return occurs in each sub-period; however, the return is only statistically significant during the 2004 to 2006 time period (at the 10 percent level). We also note that the returns in each sub-period are less than the returns available to investors by trading only on the level of short interest (see Table 6, Panel B).

The second strategy is to trade when short sellers and analysts strongly conflict about a company's prospects. The strategy is to follow the short sellers by buying firms in the portfolio with the least favorable recommendations but the lowest short interest level, while selling short firms with the best recommendations but the highest short interest. The combined return reported in Panel B is 111 monthly basis points (40 + 71 from the corner cells in Panel A), which is both

TABLE 7

Abnormal Returns to Portfolios Formed Using Recommendations and Short Interest 1994–2006

Panel A: Average Monthly Returns (Portfolio Sample Size in Parentheses) for Cross-Tabulation of Recommendation Levels and Short Interest

	Analyst Recommendation Quintiles, 1994–2006				
Short Interest Quintiles	**Least Favorable**				**Most Favorable**
Lowest Short Interest	0.40%***	0.19%	0.31%***	0.29%**	0.24%**
	(950)	(492)	(757)	(601)	(932)
	0.19%*	0.27%***	0.30%***	0.11%	0.14%*
	(799)	(691)	(820)	(691)	(788)
	0.20%*	0.20%***	0.23%***	0.03%	−0.07%
	(723)	(829)	(844)	(765)	(696)
	0.16%	0.26%***	0.00%	−0.03%	−0.18%
	(700)	(874)	(835)	(821)	(665)
Highest Short Interest	−0.12%	0.06%	−0.19%	−0.32%*	−0.71%***
	(667)	(795)	(781)	(874)	(728)

(continued on next page)

Panel B: Average Monthly Hedge Portfolio Returns

	1994–1998		1999–2003		2004–2006		Sample Period 1994–2006	
	Mean	t-stat	Mean	t-stat	Mean	t-stat	Mean	t-stat
When Short Interest and Analyst Recommendations Strongly Concur								
Firms with lowest short interest and most favorable analyst recommendation = Buy	0.01%	0.04	0.25%	1.04	0.39%	2.51**	0.24%	2.12**
Firms with highest short interest and least favorable analyst recommendation = Sell	−0.31%	−1.14	0.12%	0.24	−0.11%	−0.49	−0.12%	−0.55
Buy–Sell	0.30%	0.88	0.13%	0.22	0.50%	1.74*	0.36%	1.25
When Short Interest Strongly Conflicts with Analyst Recommendations								
Firms with lowest short interest but least favorable analyst recommendation = Buy	0.42%	1.95*	0.39%	1.45	0.25%	1.45	0.40%	2.81***
Firms with highest short interest but most favorable analyst recommendation = Sell	−0.68%	−2.68***	−0.91%	−2.95**	−0.46%	−2.11**	−0.71%	−4.27***
Buy – Sell	1.10%	2.58**	1.30%	2.62**	0.71%	2.20**	1.11%	4.09***

*, **, *** Indicate statistical significance at the $\alpha = 0.10$, 0.05, and 0.01 levels, respectively, using a two-tailed test.

n = 564,101 firm-months.

We report mean monthly abnormal returns adjusted for firm size, book-to-market ratio, and past returns (Daniel et al. 1997) for portfolios based on analyst recommendations and short interest. In parentheses, we report the average number of firms per month in each portfolio. Using the methodology in Jegadeesh and Titman (1993), we construct portfolios monthly based on the quintile rank of analyst recommendations and short interest selected in the current month and in each of the past five months (six-month holding period).

sizable and statistically significant ($t = 4.09$; $p < 0.01$). Positive returns accrue on both the long and short side, and the combined return is about double the return of 54 basis points available from trading only on the level of short interest (see Table 5, Panel B). This trading strategy also produces statistically significant returns of 110 basis points in 1994–1998, 130 basis points in 1999–2003, and 71 basis points in 2004–2006, providing evidence of the stability of this strategy over time. Note that the highest return occurs in the 1999–2003 time period when analyst recommendations are most misleading and produce a negative return of −48 monthly basis points (Table 6, Panel A).

In sum, we find that combining investment signals from analysts and short sellers yields incrementally greater future returns than trading strategies that use only one of the signals. Consistent with the results presented in Table 4, the most profitable investment strategy is when an investor trades in firms about which analysts and short sellers strongly disagree, and the investor takes the buy or sell position indicated by the short interest signal.

Robustness Tests

In this section, we report the results of several robustness tests (all untabulated). We begin by examining the sensitivity of the results to using the I/B/E/S provided consensus recommendation, rather than our self-constructed consensus. We find that our sample size increases slightly (to 86,592 firm-quarter observations) with all results qualitatively the same as those reported.

Next, we use an alternative short interest variable. Recall that for our main analyses, we use the short interest ratio (open short interest divided by shares outstanding). As an alternative deflator, we scale open short interest by the previous month's trading volume and label this variable *SIVOL*. We find that *SIVOL* is highly correlated with the short interest ratio ($\rho = 0.74$). When we regress the quintile assignment of *SIVOL* on the 11 predictive variables, we again find that all 11 variables are statistically significant in the expected direction, which is consistent with the results using the short interest ratio.

Our remaining robustness tests reexamine the profitability of the trading strategies that use information from both analysts and short sellers. First, we examine whether our results are robust to monthly rebalancing of the portfolio and using a holding period of one month. Consistent with our main results, we find that the most profitable strategy is to follow the short sellers when the signals from analysts and short sellers strongly conflict (100 monthly basis points; $t = 3.39$; $p < 0.01$). The alternative strategy, which trades when analyst and short seller signals strongly concur, remains less profitable although it improves from 36 monthly basis points ($t = 1.25$; $p = 0.21$) to 75 basis points ($t = 2.14$; $p < 0.01$).

Second, we examine whether the returns to our hedge portfolios become more profitable when we use a finer partition to form portfolios. Each month, we sort firms into deciles (instead of quintiles) based on the consensus analyst recommendation and/or the level of short interest. We intersect the recommendation and short interest deciles to produce 100 portfolios and the strategies take positions in the four most extreme portfolios using a six-month holding period. We find that the more refined stock selection yields greater hedge portfolio returns (140 monthly basis points; $t = 3.35$; $p < 0.01$).

Finally, we estimate the returns to our trading strategies using calendar-time portfolios, as an alternative method for estimating abnormal returns. Specifically, we assess the profitability of the strategies by estimating alphas from a regression of each portfolio's time-series of excess returns on the Fama-French risk factors and momentum (Fama and French 1993; Carhart 1997). We find that all hedge portfolio returns maintain statistical significance, which is consistent with the main results that calculate abnormal returns using characteristic portfolio matching (Daniel et al. 1997). We again find that the most profitable strategy is to follow the short sellers when their positions conflict with analyst recommendations.

V. CONCLUSION

We contribute new findings on the characteristics of stocks favored by short sellers, about how those characteristics differ from those used by analysts in developing buy-hold-sell recommendations, and on the value relevance of short interest data to investors. By so doing, we expand upon the results of Dechow et al. (2001) and other studies that examine information used by short sellers.[21] First, we find that analysts and short sellers use publicly available information differently. Analysts over-recommend stocks with high growth, high accruals, and low book-to-market ratios, even though prior research shows these characteristics are negatively related to future returns. In contrast, short sellers incorporate into their investment decisions the future return implications of all 11 accounting and market variables considered in this study. Second, we find that short interest provides information about future returns beyond that provided in the 11 items of information that prior research shows to be predictive of future returns. Analysts' recommendations also provide incremental information, but a negative coefficient suggests trading against the analysts. Third, based on these results, we show that a highly profitable trading strategy is one where investors trade with the short sellers when the short interest signal strongly conflicts with the consensus analyst recommendation. In fact, the value of short interest in choosing stocks to buy or sell is greater when conditioned on a conflicting consensus recommendation than when used by itself to trade stocks.

Our study contributes to the stream of academic literature documenting market inefficiencies. The debate over market efficiency has shifted from simple yes-or-no questions to issues such as the types of information incorporated into prices with a delay, the speed of price adjustment, and factors that facilitate or impede price discovery. We consider fundamental and other predictive information that prior research has shown to be incorporated into prices with a delay. We show that analysts' recommendations can impede price discovery of this information, but short selling facilitates price discovery. A frequent criticism of research documenting a delayed reaction to information is that a relation between information and future returns found to exist in the past may not recur in the future. That is, while some stocks are always misvalued, the characteristics of those stocks change over time, reducing the effectiveness of any system that places fixed weights on information. This criticism is less likely to be true for our approach, which identifies misvalued stocks based on a (strong) difference of opinion between analysts and short sellers. The characteristics of misvalued stocks can change over time, as long as short sellers and analysts disagree about valuation and the current stock price under-weights the views of short sellers. The extent of disagreement is also likely to be greatest in periods of high return volatility, and this is the type of market environment in which investors want stock-picking guidance. Consistent with this conjecture, we find the highest returns during the 1999–2003 sub-period.

Our study is timely in light of recent actions in the U.S. and other countries to further regulate short selling. For example, SEC Release No. 34-58591 (SEC 2008) requires institutional managers with at least $100 million under management to report detailed information about daily short sales in new Form SH. The rationale in Release No. 34-58591 is that "sudden and unexplained declines in the prices of securities … can give rise to questions about the underlying financial condition of an issuer, which in turn can create a crisis of confidence without a fundamental underlying basis."[22] While our evidence does not include the specific time period referred to in this quote, any new regulations are likely to extend to more normal periods of market activity. An important implication of our study is that regulations that restrict or increase the cost of short selling run the

[21] Two concurrent working papers consider aspects of short seller information use (Cao et al. 2007; Seybert and Wang 2009).

[22] In SEC Release No. 58724 (October 2, 2008), the SEC states the daily short selling information reported on Form SH will not be publicly available, in part, because it could give rise to imitative short selling.

risk of limiting a potentially important source of information for investors about future equity values. In this regard, the SEC has recently taken actions to increase the public availability of short interest positions. On March 6, 2007, the SEC approved rule changes that increase the frequency of short interest reporting from monthly to twice a month, effective September 2007. More timely reporting of short interest data to the public should further increase the role of short sellers as an information intermediary.

APPENDIX
QUANTITATIVE INVESTMENT SIGNALS

The last month of each calendar quarter is labeled quarter t. On this date, we measure our stock recommendation and short interest variables. Relative to this date, we label as quarter q the most recent fiscal quarter for which an earnings announcement is made at least two months prior to the end of quarter t and no more than four quarters prior to the end of quarter t.

Variable	Description	Calculation Details	Normative correlation with subsequent returns
SUE	Unexpected earnings	Seasonally adjusted earnings scaled by price for fiscal quarter q, as calculated by: $\frac{EPS_q - EPS_{q-4}}{Price_q}$, where EPS = earnings per share before extraordinary items (DATA#19) divided by the split adjustment factor (DATA#17) [Compustat]; and $Price$ = stock price (DATA#14) divided by the split adjustment factor (DATA#17) [Compustat].	Positive (Bernard and Thomas 1989)
TACCR	Total accruals	Earnings before extraordinary items and discontinued operations (DATA#76) minus cash flow from operations (DATA#108– DATA#78), scaled by average assets (DATA#44) as measured at the end of fiscal quarter q [Compustat]. Since Compustat reports cumulative (i.e., year-to-date) data for cash flow items, adjustments were made to arrive at total accruals for fiscal quarter q (see Collins and Hribar 2000).	Negative (Sloan 1996)
CAPEX	Capital expenditures	Rolling sum of the preceding four quarters of capital expenditures ending at fiscal quarter q divided by average total assets as calculated by: $\frac{\Sigma_{i=0}^{3} Capex_{q-i}}{(TA_q - TA_{q-4})/2}$, where $Capex$ = capital expenditures (DATA#90); and TA = total Assets (DATA#44).	Negative (Beneish et al. 2001)
MVE	Market capitalization	Natural log of the market value of equity at the end of fiscal quarter q, as calculated by DATA#14 × DATA#61 [Compustat].	Negative (Fama and French 1992)
EP	Earnings-to-price ratio	Ratio of the rolling sum of earnings over the preceding four quarters divided by price at the end of fiscal quarter q, as calculated by: $\Sigma_{i=0}^{3} \frac{EPS_{q-i}}{Price_q}$, where EPS = earnings per share before extraordinary items (DATA#19) divided by the split adjustment factor (DATA#17) [Compustat]; and $Price$ = stock price (DATA#14) divided by the split adjustment factor (DATA#17) [Compustat].	Positive (Fama and French 1992)
BTM	Book-to-market ratio	Ratio of the book value of equity to the market value of equity at the end of fiscal quarter q, as calculated by DATA#59/(DATA#14 × DATA#61) [Compustat].	Positive (Fama and French 1992)

(continued on next page)

Variable	Description	Calculation Details	Normative correlation with subsequent returns
TURN	Stock turnover	Average daily volume turnover ratio measured as the exchange-specific, percentile rank of: $\Sigma_{i=1}^{n} \frac{Daily\ Vol. / Shrout}{n}$, where *Daily Vol.* = daily stock volume [CRSP]; *Shrout* = shares outstanding [CRSP]; and n = the number of trading days for the six-month period ending on the last trading day of calendar quarter t.	Negative (Lee and Swaminathan 2000)
SG	Sales growth	Rolling sum of the preceding four quarters of sales ending at fiscal quarter q divided by the rolling sum of the preceding four quarters of sales ending on quarter $q-1$, as calculated by: $\frac{\Sigma_{i=0}^{3} Sales_{q-i}}{\Sigma_{i=0}^{3} Sales_{q-4-i}}$, where *Sales* = DATA#2 [Compustat].	Negative (Lakonishok et al. 1994)
LTG	Long-term growth forecast	Mean, consensus long-term earnings growth forecast at the end of calendar quarter t [I/B/E/S].	Negative (Lakonishok et al. 1994; La Porta 1996)
FREV	Forecast revision	Rolling sum of the preceding six-month earnings forecast revisions to price ratios, as calculated by: $\Sigma_{i=0}^{5} \frac{FEPS_{m-i} - FEPS_{m-i-1}}{Price_{m-i-1}}$, where *FEPS* = mean, consensus analyst forecast for one-year-ahead (FY1) earnings-per-share [I/B/E/S]; m = the last month of calendar quarter t; and *Price* = stock price just prior to the consensus measurement date [I/B/E/S].	Positive (Bernard and Thomas 1989; Chan et al. 1996)
MOM	Stock momentum	Buy-and-hold raw stock return for six-month period ending one month prior to the end of quarter t [CRSP].	Positive (Jegadeesh and Titman 1993)

REFERENCES

Asquith, P., and L. Meulbroek. 1995. An empirical investigation of short interest. Working paper. Massachusetts Institute of Technology.

———, M. B. Mikhail, and A. S. Au. 2005a. Information content of equity analyst reports. *Journal of Financial Economics* 75 (2): 245–282.

———, P. A. Pathak, and J. R. Ritter. 2005b. Short interest, institutional ownership, and stock returns. *Journal of Financial Economics* 78 (2): 243–276.

Barber, B. M., R. Lehavy, M. McNichols, and B. Trueman. 2001. Can investors profit from the prophets? Security analyst recommendations and stock returns. *The Journal of Finance* 56: 531–563.

———, ———, and B. Trueman. 2007. Comparing the stock recommendation performance of investment banks and independent research firms. *Journal of Financial Economics* 85 (2): 490–517.

———, ———, and ———. 2010. Ratings changes, ratings levels, and the predictive value of analysts' recommendations. *Financial Management* 39 (2): 533–553.

Barniv, R., O.-K. Hope, M. Myring, and W. B. Thomas. 2009. Do analysts practice what they preach and should investors listen? Effects of recent regulations. *The Accounting Review* 84 (4): 1015–1039.

Beneish, M. D., C. M. C. Lee, and R. L. Tarpley. 2001. Contextual fundamental analysis through the prediction of extreme returns. *Review of Accounting Studies* 6 (2–3): 165–189.

Bernard, V. L., and J. K. Thomas. 1989. Post-earnings-announcement drift: Delayed price response or risk premium? *Journal of Accounting Research* 27 (Supplement): 1–36.

Boehmer, E., Z. R. Huszar, and B. D. Jordan. 2010. The good news in short interest. *Journal of Financial Economics* 96 (1): 80–97.

———, C. M. Jones, and X. Zhang. 2008. Which shorts are informed? *The Journal of Finance* 63 (2): 491–527.

——— 2009. Shackling short sellers: The 2008 shorting ban. Working paper. University of Oregon.

Bradshaw, M. T. 2004. How do analysts use their earnings forecasts in generating stock recommendations? *The Accounting Review* 79 (1): 25–50.

———, S. A. Richardson, and R. G. Sloan. 2001. Do analysts and auditors use information in accruals? *Journal of Accounting Research* 39 (1): 45–74.

Cao, B., D. Dhaliwal, A. C. Kolasinski, and A. V. Reed. 2007. Bears and numbers: Investigating how short sellers exploit and affect earnings-based pricing anomalies. Working paper. University of Washington.

Carhart, M. M. 1997. On persistence in mutual fund performance. *The Journal of Finance* 52 (1): 57–82.

Chan, L. K. C., N. Jegadeesh, and J. Lakonishok. 1996. Momentum strategies. *The Journal of Finance* 51 (5): 1681–1713.

Collins, D. W., and P. Hribar. 2002. Errors in estimating accruals: Implications for empirical research. *Journal of Accounting Research* 40 (1): 105–134.

Daniel, K., M. Grinblatt, S. Titman, and R. Wermers. 1997. Measuring mutual fund performance with characteristic-based benchmarks. *The Journal of Finance* 52 (3): 1035–1058.

Dechow, P. M., A. P. Hutton, L. Meulbroek, and R. G. Sloan. 2001. Short-sellers, fundamental analysis, and stock returns. *Journal of Financial Economics* 61 (1): 77–106.

Demirakos, E., N. Strong, and M. Walker. 2004. What valuation models do analysts use? *Accounting Horizons* 18 (4): 221–240.

Desai, H., K. Ramesh, S. R. Thiagarajan, and B. V. Balachandran. 2002. An investigation of the informational role of short interest in the NASDAQ market. *The Journal of Finance* 57 (5): 2263–2287.

Diamond, D. W., and R. E. Verrecchia. 1987. Constraints on short-selling and asset price adjustment to private information. *Journal of Financial Economics* 18 (2): 277–311.

Fama, E. F., and K. R. French. 1992. The cross-section of expected stock returns. *The Journal of Finance* 47: 427–465.

——— 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33 (1): 3–56.

Frankel, R., and C. M. C. Lee. 1998. Accounting valuation, market expectation, and cross-sectional stock returns. *Journal of Accounting and Economics* 25: 283–319.

Gow, I. D., G. Ormazabal, and D. Taylor. 2010. Correcting for cross-sectional and time-series dependence in accounting research. *The Accounting Review* 85 (2): 483–512.

Jegadeesh, N., and S. Titman. 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance* 48 (1): 65–91.

———, J. Kim, S. D. Krische, and C. M. C. Lee. 2004. Analyzing the analysts: When do recommendations add value? *Journal of Finance* 59 (3): 1083–1124.

Lakonishok, J., A. Shleifer, and R. W. Vishny. 1994. Contrarian investment, extrapolation, and risk. *The Journal of Finance* 49: 1541–1578.

La Porta, R. 1996. Expectations and the cross-section of stock returns. *The Journal of Finance* 51: 1715–1742.

Lee, C. M. C., and B. Swaminathan. 2000. Price momentum and trading volume. *The Journal of Finance* 55 (5): 2017–2070.

Lin, H.-W., and M. F. McNichols. 1998. Underwriting relationships, analysts' earnings forecasts and investment recommendations. *Journal of Accounting and Economics* 25 (1): 101–127.

Ljungqvist, A., C. Malloy, and F. Marston. 2009. Rewriting history. *The Journal of Finance* 64 (4): 1935–1960.

Malmendier, U., and D. Shanthikumar. 2007. Are small investors naive about incentives? *Journal of Financial Economics* 85: 457–489.

Miller, E. M. 1977. Risk, uncertainty, and divergence of opinion. *The Journal of Finance* 32: 1151–1168.

Petersen, M. A. 2009. Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies* 22 (1): 435–480.

Piotroski, J. D. 2000. Value investing: The use of historical financial statement information to separate winners from losers. *Journal of Accounting Research* 38 (Supplement): 1–41.

Pownall, G., and P. Simko. 2005. The information intermediary role of short sellers. *The Accounting Review* 80 (3): 941–966.

Securities and Exchange Commission (SEC). 1999. *Short Sales*. SEC Concept Release No. 34-42037. File S7-24-99. Washington, D.C.: Government Printing Office.

———. 2008. Release No. 34-58591. Available at: http://www.sec.gov/rules/other/2008/34-58591.pdf.

Seybert, N., and S. Wang. 2009. False prophets: How analysts lead short sellers astray during periods of high investor sentiment. Working paper. The University of Texas at Austin.

Sloan, R. G. 1996. Do stock prices fully reflect information in accruals and cash flows about future earnings? *The Accounting Review* 71 (3): 289–315.

Swanson, E. P., L. Rees, and L. F. Juarez-Valdes. 2003. The contribution of fundamental analysis after a currency devaluation. *The Accounting Review* 78 (3): 875–903.

Thompson Financial. 2009. *Methodology for Estimates*. Available at: http://www.library.kent.edu/files/TF_Methodology_for_Estimates_October_2009.pdf

Womack, K. 1996. Do brokerage analysts' recommendations have investment value? *The Journal of Finance* 51: 137–167.

THE ACCOUNTING REVIEW
Vol. 86, No. 1
2011
pp. 131–153

American Accounting Association
DOI: 10.2308/accr.00000007

Incentive Compensation and Promotion-Based Incentives of Mid-Level Managers: Evidence from a Multinational Corporation

Merle Ederhof
University of Michigan

ABSTRACT: This study re-examines the hypothesis that explicit, compensation-based incentives of mid-level managers are adjusted to the level of implicit incentives provided by the possibility of moving to higher-level positions. Using compensation data from a large multinational corporation, I find that, after controlling for the position's scope and level of accountability, bonus-based incentives are stronger for managers who (1) have fewer organizational levels left to climb, (2) face weaker implicit incentives from getting promoted to the next level, and (3) face weaker implicit incentives from getting promoted to the top of the organization. The findings are consistent with the notion that implicit incentives are taken into consideration in the design of explicit incentive contracts. In particular, the results support the prediction that explicit incentives are optimally stronger in situations with weaker implicit incentives.

Keywords: *implicit incentives; incentive compensation; promotions; career concerns.*

Data Availability: *The data used in this study are proprietary.*

I. INTRODUCTION

The theoretical literature has argued that managers who face weaker promotion-based implicit incentives should optimally receive stronger explicit variable-pay-based incentives (Gibbons and Murphy 1992; Gibbs 1995). Despite well-developed theoretical arguments, empirical studies have had limited success providing evidence that implicit, promotion-based incentives are taken into consideration in the design of explicit incentive compensation contracts.[1] The present study revisits this hypothesis.

I thank Shannon Anderson (the editor), John Core, Ian Gow, Lisa Kahn, Brian Mittendorf, Madhav Rajan, Suraj Srinivasan, two anonymous reviewers, and seminar participants at the Yale Fall Conference, the NYU/Yale joint seminar, Michigan State University, University of Alberta, University of Michigan, University of Connecticut, and the AAA Annual Meeting. I also thank Bjoern Anton for excellent research assistance.

Editor's note: Accepted by Shannon Anderson.

[1] Gibbs (1995) analyzes the explicit incentives of employees who have been passed over for promotion; his findings do not indicate significant differences to the explicit incentives of employees who have not been passed over for promotion. Ortin-Angel and Salas-Fumas (1998) show that managers at higher organizational levels have stronger explicit incentives. However, their findings could be attributable to the fact that managers at higher levels are in positions that have more decision-making authority.

Submitted: June 2009
Accepted: June 2010
Published Online: January 2011

I analyze a sample of mid-level managers who can be directly compared with respect to their positions, but who occupy different positions in their respective hierarchies and who face different promotion possibilities and rewards upon getting promoted. Thus, the setting provides an opportunity to observe variation in the strength of the implicit, promotion-based incentives while controlling for many confounding factors. In particular, the analyses in this study are based on compensation data from a large multinational corporation that operates in over 100 countries around the world. The company is organized around five main divisions, which in turn are comprised of 30 subdivisions. Each of the divisions and subdivisions is represented in many different countries. This matrix-like organizational structure allows me to directly compare positions across countries.

I find that the explicit incentives provided by the company's bonus plan are stronger for managers who are positioned at higher organizational levels, face weaker implicit incentives from getting promoted to the next level, and face weaker implicit incentives from getting promoted to the top of the organization, after controlling for the position's scope and level of accountability. These findings are consistent with the theoretical argument that implicit incentives should be taken into consideration in designing explicit incentive contracts. More precisely, the evidence presented here supports the prediction that explicit incentives are optimally stronger in situations where implicit incentives are weaker (Gibbons and Murphy 1992; Gibbs 1995).

This study makes several contributions to the incentive compensation literature. First, I extend prior empirical work in accounting that examines the incentive intensity of mid-level managers (Baiman et al. 1995; Nagar 2002) to consider the role of implicit promotion-based incentives, thereby following the call in Bushman and Smith (2001) for more research on the interactions between incentive contracting and other organizational features. In particular, the present study documents that the intensity of explicit incentives is higher in situations that pose weaker implicit, promotion-based incentives.

Second, this study contributes to prior studies that look at the interplay between explicit incentive contracts and implicit incentives arising from the possibility of career advancement (e.g., Kahn and Sherer 1990; Gibbons and Murphy 1992). Those studies primarily focus on the length of the manager's career horizon as the measure of the strength of implicit incentives. This study adds to that literature by analyzing an innovative setting that provides the opportunity to isolate the strength of implicit, promotion-based incentives. The analyses in this study provide additional empirical evidence on the importance of career-based incentives in contract design.

Arguably most closely related to this study, Gibbs (1995) analyzes the compensation-based incentives of employees who have been passed over for promotion. The author does not find significant differences between the explicit incentives of employees who have been passed over for promotion and those of employees who have not been passed over for promotion. Gibbs (1995) hypothesizes that the lack of evidence could be attributable to a centrally administered incentive scheme, which may not allow for variation at the employee level. In contrast to the setting analyzed by Gibbs (1995), the empirical setting in this study provides the following advantages. First, the data set used in this study includes explicit information on the parameters of the incentive contracts, such as the expected bonuses, which allows for a more precise measurement of the strength of the bonus-based incentives. Second, although Gibbs (1995) and the present study both analyze data from a single company, in my empirical setting, the explicit incentives are determined by the individual country organizations. Thus, my analysis circumvents the issue of a centrally administered incentive plan.

The ability to generalize the results of this study is limited by the analysis of a single firm. However, the research site chosen for this study offers several advantages for empirical investigation of the role of implicit, promotion-based incentives in incentive contracts provided to mid-level managers. First, the organizational structure of the company, combined with the company's

job-rating project, allows for a direct comparison of managers in different countries. Second, the compensation practices in the different countries reflect the local labor markets and, thus, are not specific to the company that is studied. Finally, the incentives provided to the managers through the company's bonus plan are not based on company-wide guidelines, but are decided by the individual country organizations.

The next section develops the hypotheses. Section III describes the research site, the sample and the measures used in the empirical analyses. Section IV discusses the research design and the empirical results. Section V provides a summary and conclusion.

II. HYPOTHESIS DEVELOPMENT

This study analyzes the incentives of mid-level managers who work in a corporate hierarchy. In particular, I investigate whether explicit incentives that are provided by variable-pay-based schemes are adjusted, based on the level of implicit incentives that are provided by the possibility of moving to higher-level positions in the organization.[2]

The argument that explicit, variable-pay-based incentives are optimally stronger in situations where implicit, promotion-based incentives are weaker has been formalized in career concerns models that allow for the presence of explicit incentive contracts (Gibbons and Murphy 1992; Gibbs 1995).

In particular, Gibbons and Murphy (1992) investigate optimal (explicit) incentive contracts when the agent faces implicit incentives from the possibility of career advancement in a competitive external labor market. The analysis in Gibbons and Murphy (1992) is based on a multiperiod model with a single performance measure, which is a function of the agent's innate ability and the agent's effort that is provided during the period. The performance metric is used in the explicit incentive contract and is also used by the external labor market to update beliefs about the agent's ability. The compensation that is offered to the agent in the second period in the competitive labor market is increasing in the market's assessment of his/her ability. Implicit incentives arise in this setting because ability and effort cannot be fully separated and the agent has the incentive to increase effort to influence the labor market's beliefs about his/her ability.[3] The implicit incentives provided by career concerns are stronger when future compensation is more valuable to the agent as is the case when the agent is further away from retirement. Given this setup, the analysis shows that the optimal explicit incentives provided by the agent's compensation scheme are decreasing in the implicit incentives provided by career concerns.

Although Gibbons and Murphy (1992) examine implicit incentives that arise in a competitive external labor market, their analysis can also be interpreted in the light of an internal labor market, as it is discussed by the authors on pages 469–470. In particular, career concerns also arise in an internal labor market if the employee's supervisor cannot perfectly distinguish between the employee's ability and his/her effort. The authors argue that in an internal labor market setting explicit incentives should be strongest for workers with weak promotion-based incentives such as workers at the top of the corporate hierarchy.

Consistent with the result in Gibbons and Murphy (1992), Gibbs (1995), using a single-period model, shows that the explicit incentives provided by a compensation scheme are optimally stronger when the implicit incentives provided by the possibility of being promoted are weaker. In Gibbs' (1995) model, the promotion decision depends on the outcome of the performance measure, which again provides incentives for the agent to increase his/her effort.

---

2 Promotion-based incentives are one source of implicit incentives. Other forms of implicit incentives include incentives that are based on non-contractible performance measures (see, e.g., Ederhof 2010).

3 In equilibrium, the market's conjecture about the worker's ability is correct but the agent will exert higher effort than in the absence of career concerns because the market discounts his/her effort.

An important feature of the models in Gibbons and Murphy (1992) and Gibbs (1995) is that the strength of the implicit incentives is not a choice variable for the principal. In other words, both models characterize how the optimal explicit incentive contract should be designed for a given level of implicit incentives. This is in contrast to Lazear and Rosen (1981) and Rosen (1986) who analyze how the principal should optimally choose the compensation structure across hierarchical levels in order to optimize the resulting implicit incentives. It seems reasonable to assume that, in my empirical setting, the implicit incentives provided by the possibility of promotion are determined exogenously with respect to the manager's compensation scheme. The company investigated in this study is organized along five main divisions with operations in many different countries. The purpose of the local units of a division in the different countries is to implement the global strategy of the division in the individual countries. The organizational structure of the local units is largely standardized. For example, all worldwide local units of a division are organized around a local unit manager whose authority and responsibilities follow worldwide guidelines. Thus, it seems unlikely that the company adapts its organizational form in a given country to the local compensation structure. In other words, it seems unlikely that the company structures its organizational form around the local labor market in order to optimize the implicit incentives resulting from the possibility of career advancement. Moreover, the compensation paid to the local managers is dictated by the local labor market conditions (Gibbs 1995).

Broadly speaking, the result of the analyses in Gibbons and Murphy (1992) and Gibbs (1995) is that optimal explicit incentives are decreasing in the strength of the implicit incentives provided by the possibility of career advancement. With respect to the setting of a corporate hierarchy, the strength of the implicit incentives is determined by the extent to which additional effort changes the probability of getting promoted and the "prize" that the manager is awarded upon promotion. The prize of getting promoted, in turn, is comprised of the immediate increase in compensation and the option value of being eligible for future rewards deriving from further promotions (Rosen 1986; Gibbs 1995).

An important determinant of the strength of the implicit incentives that a manager faces is his/her hierarchical position, because managers who are closer to the top of their organization have "truncated" promotion paths. Thus, high-rank managers are expected to have stronger explicit incentives. However, managers who are at higher organizational levels are likely to have stronger explicit incentives due to their job characteristics. In particular, managers at higher hierarchical levels are likely to have higher marginal productivities with respect to their effort (Baker and Hall 2004). Moreover, managers at high organizational ranks are likely to have more decision-making authority and to have a larger span of control (Prendergast 2002; Nagar 2002; Wulf 2007). In order to isolate the strength of implicit incentives, I use various control variables to capture variation in such job characteristics in the empirical analysis. The hypothesis can be stated as follows (expressed in the alternative form):

Hypothesis: The explicit incentives provided by the variable-pay scheme are decreasing in the strength of the promotion-based implicit incentives that the manager faces, *ceteris paribus*.

III. RESEARCH SETTING AND MEASURES

Research Site

I test the hypothesis using data from a large multinational engineering corporation that operates in approximately 100 countries. The company primarily sells technology to utility and industry customers. The company's operations are organized into five main divisions, which in turn are

comprised of 30 subdivisions.[4] The divisions are the central building blocks of the organization; they run the business lines from R&D to sales and they have primary P&L responsibility. For each of the divisions and subdivisions, a manager is responsible for the unit's worldwide operations. For each country in which a unit operates, a local manager is responsible for the unit's local operations. In addition, the countries have some infrastructure in the form of a country management team and support functions such as human resources, finance, legal, and communication.

I test the hypothesis by comparing the incentives of the local managers in the different countries. In particular, the sample is comprised of all local managers who are in positions that are part of their local organizational "ladders," which go all the way to the top of the organizations. In other words, the sample includes managers who hold positions that make them eligible to eventually rise to the top of their respective organization. The sample excludes positions like IT and legal, which constitute support functions at the company and for which the promotion possibilities of the positions' holders are limited.[5]

The Company's Job-Rating Project

In 2005, with the help of a consulting firm, the company started a project of assigning numerical ratings to the top positions in the company. The company initiated the project for the following reasons. First, and important for this study, the company wanted to generate a picture of the organization's hierarchy that reflects managers' promotion paths and that is independent of existing job titles.[6] In interviews with the author, the company contact emphasized that employees often have mental models of the company hierarchy that is based on job titles that do not correspond to the true hierarchy of the organization. The company hopes that the job-rating project will improve the promotion process, in that it will facilitate, for example, promoting people to jobs that have a higher rating but that have the identical title as the employee's current position.

Second, the company wanted to gain an understanding of how comparable positions are compensated in the different countries. In interviews with the author, the company contact emphasized that, to that end, a key aspect of the job-rating project was that the ratings were not influenced by the current position holder's compensation. Moreover, in order to facilitate comparability across countries, the ratings were assigned based on a standardized scheme, which the consulting company developed based on an initial pilot study.

The ratings are assigned to a position, and not to the person currently holding the position. Thus, the ratings are independent of the current manager's performance. The ratings are assigned based on a combination of factors primarily capturing the position's scope and level of accountability. In particular, a position's scope is captured by the number of employees that a manager oversees and the revenue figure for which the manager is responsible. The level of accountability is largely captured by the decision-rights that are in the manager's hand.

In interviews with the author, the company contact emphasized that promotions occur from one rating category to the next. In 2008, the company completed rating all positions that fall into the top ten rating categories. The position of the chief executive officer is assigned a rating of 1; an example of a job with a rating of 10 is a manager who is in charge of the operations of a local subdivision with revenue of $22 million and 40 employees.

---

4 Subsequently, I collectively refer to divisions and subdivisions as "units."

5 The classification of positions into "organizational ladder" and "support functions" is based on extensive interviews with the company contact.

6 The notion that the organization's hierarchy is based on promotion paths is consistent, for example, with the way the hierarchy in Baker et al. (1994) and Gibbs (1995) is defined.

The Company's Incentive System

Aside from a fixed salary, all managers in the sample are eligible for a bonus payment. The company has guidelines that pertain to all worldwide participants in the bonus plan. In particular, the performance measures used in the bonus plan, the weight that is placed on firm-wide versus divisional measures, and aspects of the pay-performance relation must follow the company's guidelines. In contrast, and important for this study, the company does not have worldwide guidelines with respect to participants' expected bonuses—i.e., the compensation that is paid out in the form of a bonus when the performance meets expectations. In other words, the company does not have guidelines with respect to the percentage of a participant's cash compensation that is variable. The level of the expected bonus, which is typically expressed as a percentage of base salary, is determined by the respective country management. Some countries interpret the bonus plan as a guaranteed 13th-month salary for all participants. These countries are excluded from the analysis.

In addition to the bonus plan, the company awards stock options to the top employees of the organization who hold positions that are largely in the top seven rating categories. In contrast to the company's bonus plan, all aspects of the stock option plan are centrally administered by the company's headquarters and follow worldwide company guidelines. In awarding stock options in 2008, the company made use of the newly available job ratings. In particular, the number of options awarded to managers in a given job category is fixed for a given job-rating category.

Sample and Measures

The analyses in this study are based on a data set that contains information for 1,151 managers in 14 countries that have been assigned ratings between 2 and 10. The data are largely for the year 2008. Table 1 shows the distribution of the sample across different job-rating categories and countries.

Table 2 and Figure 1 summarize compensation levels across organizational levels for the different countries. Specifically, the table and graph show how the median total cash compensation, which is calculated by summing base salary and expected bonus, varies across job categories in the different countries.

The compensation figures are expressed relative to the median compensation levels at job-rating category 10 in the respective country, which are normalized to 1.00.[7] As one would expect, for the most part, the values of the compensation ratios are increasing as managers move to higher-level job categories. Visual inspection of Figure 1 suggests that the overall pay structures can be characterized as convex, which is consistent with findings in prior studies and the prediction from tournament theory (e.g., Lambert et al. 1993; Rosen 1986). Table 2 and Figure 1 also indicate substantial variation in the pay structures across countries. For example, the ratio of median pay levels in category 5 to category 10 is approximately 2 in Germany, but more than 4 in the United Kingdom.

Explicit Incentives

Testing the hypothesis requires a measure of the explicit incentives provided by the compensation scheme. As discussed above, stock options are awarded uniformly across countries and the number of options awarded is fixed for a given job-rating category. Therefore, I focus attention on the company's bonus plan. Conceptually, the strength of the incentives provided by a bonus plan is reflected in how much the agent's compensation increases when s/he increases his/her effort. As discussed above, the company has worldwide guidelines with respect to certain parameters of the

---

[7] Confidentiality reasons preclude me from reporting dollar amounts.

TABLE 1
Distribution of the Sample across Different Rating Categories and Countries

Country	Rating 1	2	3	4	5	6	7	8	9	10	Total
Sweden	0	1	0	1	5	9	15	40	78	99	248
Finland	0	0	1	0	0	5	8	12	34	37	97
Germany	0	0	1	0	1	3	9	30	32	18	94
Italy	0	0	1	1	3	5	10	21	23	28	92
Switzerland	0	0	0	1	3	6	6	24	42	92	174
Norway	0	0	0	0	2	4	3	9	15	18	51
Poland	0	0	0	0	1	0	0	6	5	11	23
Russia	0	0	0	0	1	1	3	6	5	12	28
Spain	0	0	0	0	2	2	1	5	12	21	43
United Kingdom	0	0	0	0	1	2	7	8	17	12	47
United States	0	0	0	0	4	17	26	40	63	74	224
Denmark	0	0	0	0	0	1	1	3	5	1	11
New Zealand	0	0	0	0	0	1	0	1	3	1	6
Turkey	0	0	0	0	0	1	1	4	2	5	13
Total	0	1	3	3	23	57	90	209	336	429	1,151

TABLE 2

Median Pay Levels across Job-Rating Categories for the Different Countries (expressed relative to the median pay level for category 10 in the respective country)[a]

Job-Rating Category	Sweden	Finland	Germany	Italy	Switzerland	Norway	Poland
2	5.93	—	—	—	—	—	—
3	—	5.69	3.26	—	—	—	—
4	3.81	—	—	2.77	3.40	—	—
5	2.80	—	2.05	2.42	2.34	2.50	3.76
6	1.98	2.45	1.64	2.80	2.06	1.66	—
7	1.65	1.65	1.46	2.29	1.80	1.69	—
8	1.25	1.31	1.15	1.43	1.29	1.40	1.69
9	1.08	1.23	1.04	1.27	1.07	1.12	1.24
10	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Job-Rating Category	Russia	Spain	United Kingdom	United States	Denmark	New Zealand	Turkey
2	—	—	—	—	—	—	—
3	—	—	—	—	—	—	—
4	—	—	—	—	—	—	—
5	3.81	2.89	4.35	3.16	—	—	—
6	3.28	2.06	3.24	1.83	3.38	2.88	4.39
7	1.94	1.70	1.77	1.63	1.16	—	4.31
8	1.60	1.48	1.54	1.38	1.10	1.66	2.71
9	1.38	1.15	1.08	1.11	0.79	1.24	1.57
10	1.00	1.00	1.00	1.00	1.00	1.00	1.00

[a] Pay is the sum of the manager's base salary and his/her expected bonus. Expatriates are excluded from the statistics.

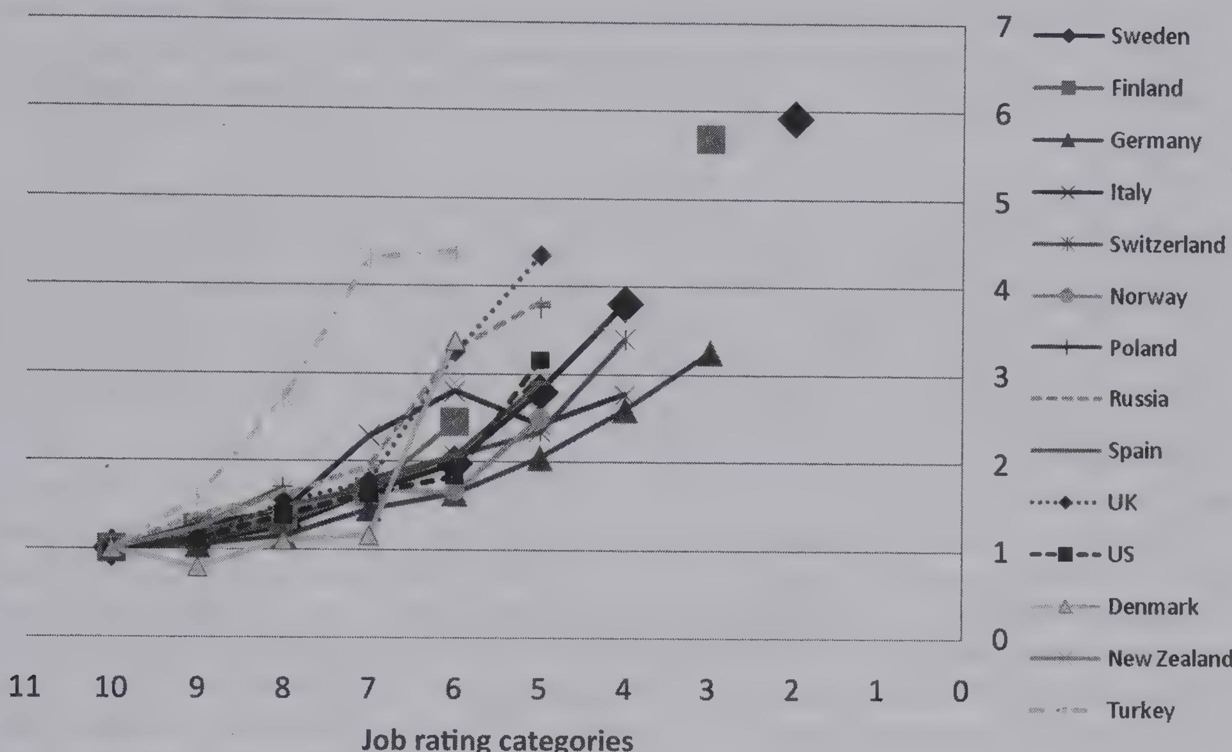
bonus plan. In particular, the payoff function is linear and an increase in performance leads to a fixed percentage increase in bonus payout. For example, the bonus payout for the maximum performance is twice as much as the bonus payout for the target performance for all plan participants. Therefore, variation in the expected bonus reflects variation in the strength of the incentives provided by the bonus plan. Thus, I measure the strength of the explicit incentives provided by the bonus scheme using the ratio of the target bonus to base salary (*TB*) (Indjejikian and Nanda 2002).

Implicit Incentives

The hypothesis predicts that managers who face stronger promotion-based implicit incentives have lower variable-pay-based explicit incentives. As discussed above, the strength of the implicit incentives is determined by the extent to which additional effort changes the probability of getting promoted and the "prize" that the manager is awarded upon promotion. The prize of getting promoted, in turn, is comprised of the immediate increase in compensation and the option value of being eligible for future rewards deriving from further promotions (Rosen 1986; Gibbs 1995). I use different measures to capture the strength of the implicit incentives.

First, as argued above, managers at higher hierarchical positions in their respective organizations have truncated promotion paths and are thus expected to have lower implicit incentives. In order to capture a manager's hierarchical position in his/her country organization, I include the job

FIGURE 1
(Median Cash Pay at Respective Level/Median Cash Pay at Level 10)[a]



[a] Expatriates are excluded from the statistics.

rating of the highest-ranking manager in the respective country, *HIGHESTPOSITION*, in the analysis.

As described above, the strength of the promotion-based implicit incentives is a function of the prize that the manager receives upon promotion and the extent to which additional effort changes the probability of getting promoted. I develop two additional measures to capture those features.

It is unobservable how additional effort changes the probability of getting promoted. However, Gibbs (1995, 1996) has shown that the derivative of the probability of getting promoted with respect to the agent's effort is increasing in the promotion probability as long as the promotion probability is below one-half (also see Campbell 2008). Interviews with the company contact confirmed that it can reasonably be assumed that, for the managers in my sample, the implicit incentives are increasing in the promotion probabilities, as the promotion rates at the company are sufficiently low.

Data limitations prevent the direct calculation of the promotion probabilities in the individual job categories. However, the company contact emphasized that dismissals and demotions are fairly rare in this company (also see Gibbs 1995). Thus, I employ the median tenure at the individual job levels as a proxy for promotion probabilities. More precisely, I use the inverse of the median tenure in the manager's current job-rating category to capture promotion possibilities.

Focusing on the manager's immediate promotion possibility, I construct a measure that takes the compensation differential between the manager's current job and the next-higher job-rating category, and the manager's chance of being promoted to that next level into consideration. Specifically, I compute the ratio of the median expected cash compensation for the next-higher job-rating category to manager's expected cash compensation (in his/her current job), both for the manager's respective country. I then multiply that ratio by the inverse of the median tenure in the manager's current job-rating category in the respective country. The measure is denoted by *IMPLICITNEXT*.

Similarly, consistent with Rosen (1986), I develop a measure that takes the median tenure and compensation level at the manager's current job and at each higher-level category in the respective organization into consideration. Specifically, I calculate the sum of the discounted compensation differentials between the manager's current job and the top of the respective organization. The compensation differential between two levels is discounted by the cumulative probability of being promoted to the respective level, where the probability of promotion is again proxied for by the inverse of the median tenure. I denote this measure as *IMPLICITTOP*.

Control Variables

Incentive compensation practices vary systematically across countries. As discussed in more detail in the research design section, I employ a measure that captures the general incentive intensity in the respective country. Specifically, I use a measure that is provided by the company's consulting company that captures the median incentive intensity for mid-level managers for a large number of industrial companies in the different countries.[8]

Prior literature has argued that division managers who have more decision-making authority receive more incentive-based pay because the potential for misuse is stronger if managers have more authority (Prendergast 2002; Nagar 2002; Wulf 2007). I use the job rating assigned to the manager's position, *CATEGORY*, as a proxy for the manager's decision-making authority. As described above, the ratings are assigned based on factors that primarily capture the positions' scope and level of accountability, an important determinant of which is the manager's level of decision-making authority.

A standard result in agency theory is that managers who have a higher marginal productivity with respect to effort should optimally have stronger explicit incentives (Baker and Hall 2004). In order to control for differences in managers' marginal productivities, following Wulf (2007), I include a measure of the relative importance of the manager's unit in the analyses. In particular, I measure a unit's relative importance by the ratio of the sales of the unit that the manager is affiliated with to the total sales in the respective country (*RSALES*).[9]

Theory suggests that more noise in the performance measures increases the risk that the manager is exposed to and the prediction is that incentives are lowered when the risk exposure is higher (Holmstrom 1979).[10] As described above, the performance measures that are used in the bonus plans follow company guidelines in the sense that, worldwide, managers in the same position have the same performance measures in their bonus plan. Thus, managers in similar positions in different countries could be exposed to different levels of noise because they operate in different environments. Due to data limitations, I rely on the measure capturing the general incentive intensity in the different countries in order to capture differences in noise levels.

I also include the expected growth in sales in the regression analysis. I measure expected growth in sales by the ratio of budgeted sales for 2008 to the actual sales number for 2007

---

8 I do not report descriptive statistics on this measure due to confidentiality reasons.
9 Also see Baiman et al. (1995).
10 Also see Prendergast (2000) for arguments why the relationship between uncertainty and incentives could be positive.

(*SALESGROWTH*). This measure can be interpreted as a proxy for the firm's investment opportunities. Prior literature has argued that firms with greater growth opportunities employ compensation contracts with greater incentive intensity. Smith and Watts (1992) argue that the observability of managers' actions decreases with the firm's growth opportunities. In contrast, the actions of managers in low-growth firms are argued to be more observable because these actions are largely focused on the maintenance and supervision of existing assets (Gaver and Gaver 1993; Holthausen et al. 1995).[11]

Agency theory has also argued that the optimal incentive intensity depends on the level of monitoring (e.g., Jensen and Meckling 1976; Prendergast 2002; Liang et al. 2008). In particular, Jensen and Meckling (1976) argue that incentive contracting and monitoring are alternative solutions to the moral hazard problem. As discussed above, the company is organized around five main divisions, which in turn are comprised of 30 subdivisions. The individual country organizations are structured around a country management team. Moreover, the local divisions and subdivisions are led by local division and subdivision managers. It seems reasonable to expect that direct monitoring of the actions of managers who are at the top of an organization is more difficult than monitoring the actions of lower-level managers. In order to capture such differences, I include an indicator variable, *TOPMANAGER*, indicating whether a manager is a highest-ranking manager in his/her country. Moreover, it seems plausible that local division managers do not receive as much monitoring as managers who are at lower organizational levels. Specifically, local division managers lead the respective business lines in their country. Thus, they are the highest-ranking managers in their respective fields of expertise. In order to capture potential differences in the monitoring of local division managers, I include an additional indicator variable, *DIVMANAGER*, in the tests. All measures are defined in Table 3.

Descriptive Statistics

Of the 1,151 managers in the sample, 18 are expatriates. In the company, expatriates are compensated based on the norms in their home country with certain adjustments, such as for hardship and allowance. Thus, expatriates are excluded from the analyses. With the exception of Italy, the expatriates hold lower-level positions, with ratings between 6 and 10. In Italy, the highest-ranking manager is an expatriate, which precludes calculation of the variable *IMPLICIT-TOP* for the observations from that country.

Table 4 provides descriptive statistics for the variables used in the analysis, calculating such statistics for each job-rating category using the pooled sample across all countries.

The mean (median) values for the dependent variable in the analyses, *TB*, which captures the ratio of (target bonus/salary), are 0.18 (0.14) for managers at level 10 and 0.39 (0.39) for managers who hold positions in job-rating category 3. Overall, the descriptive statistics indicate an increasing trend of *TB* as one moves to higher-level positions, which is consistent with the expectation that managers with more decision-making authority receive more incentive-based pay. However, the mean and median values for *TB* are fairly constant for job-rating categories 2 through 5. One issue that should be kept in mind is that the job-rating categories are not distributed evenly across countries and that there are country-specific differences in the level of incentive compensation. The finding that the mean and median values for *TB* are fairly constant for job-rating categories 2 through 5 could also be influenced by the fact that there is no one-to-one mapping between job-rating categories and hierarchical levels. Specifically, each of the categories 2 through 5

---

[11] Lambert and Larcker (1987) explain why the relationship between expected sales growth and the strength of the incentives could also be negative.

TABLE 3
Measures

Measure Reflecting the Strength of Explicit Incentives	
TB_{ij}	(target bonus$_i$/base salary$_i$); both for manager i.
Measures Reflecting the Strength of the Implicit Promotion-Based Incentives	
$HIGHESTPOSITION_j$	job-rating category of the highest-ranking manager in country j.
$IMPLICITNEXT_{ij}$	[(median expected cash pay$_{jt-1}$/expected cash pay$_i$) $*$ (1/median tenure$_{jt}$)]: the ratio of the median cash pay (base salary + expected bonus) for job-rating category $t-1$ in country j to the manager's expected cash pay in his/her current position.
$IMPLICITTOP_{ij}$	$(\text{median expected cash pay}_{jt-1} / \text{expected cash pay}_i) * (1 / \text{median tenure}_{jt}) + \sum_{a=1}^{t-1} [(\text{median expected cash pay}_{ja-1} / \text{median expected cash pay}_{ja}) * \prod_{b=0}^{t-a} (1 / \text{median tenure}_{ja+b})]$: sum of the discounted compensation differentials between the manager's current job-rating category t and the top of the respective country organization j.
Control Variables	
$CATEGORY_{ij}$	job rating assigned to manager i's position.
$RSALES_{ij}$	(sales of unit k/total sales of country j that unit k is located in).
$SALESGROWTH_{ij}$	(budgeted sales for 2008$_k$/actual sales for 2007$_k$); both for unit k in country j.
$TOPMANAGER_{ij}$	indicator variable that is equal to 1 if *CATEGORY* indicates that the manager in job-rating category t is a highest-ranking manager in country j, 0 otherwise.
$DIVMANAGER_{ij}$	indicator variable that is equal to 1 if manager i is in charge of a local division, 0 otherwise.

includes a high percentage of managers who are highest-ranking in their respective country organization.

The descriptive statistics for *IMPLICITNEXT* suggest that the implicit incentives provided by the prospect of solely moving up one job category are largely increasing in the hierarchical level. Specifically, the mean (median) values for *IMPLICITNEXT* are 0.38 (0.37) for managers at level 10 and 0.61 (0.63) for managers who hold positions in job-rating category 5. This trend is consistent with the convex pay structures documented in Table 2 and Figure 1.

The variable *IMPLICITTOP* captures the implicit incentives provided by the possibility of moving to the top of the respective organization. Although the descriptive statistics largely indicate an upward trend with respective to the mean levels as one moves up the organizational hierarchy, they do not suggest a strong trend with respect to the median values. Specifically, with the exception of the figure for job-rating category 6, the median values of *IMPLICITTOP* are around 0.60 to 0.65 across categories. This finding is consistent with the notion in tournament theory that convex pay structures result in constant implicit incentives throughout the hierarchy (Rosen 1986). Broadly speaking, the intuition is as follows. On the one hand, managers who are further down in the hierarchy face a higher number of organizational levels that they could potentially climb, which increases the number of terms that comprise the option value, which

TABLE 4
Descriptive Statistics[a]

	n	Mean	Std.	Q1	Median	Q3
CATEGORY = 2						
TB	1	0.38	NA	0.38	0.38	0.38
IMPLICITNEXT	NA	NA	NA	NA	NA	NA
IMPLICITTOP	NA	NA	NA	NA	NA	NA
TOPMANAGER	1	1.00	NA	1.00	1.00	1.00
DIVMANAGER	1	1.00	NA	1.00	1.00	1.00
RSALES	1	0.92	NA	0.92	0.92	0.92
SALESGROWTH	1	1.19	NA	1.19	1.19	1.19
CATEGORY = 3						
TB	2	0.39	0.02	0.38	0.39	0.40
IMPLICITNEXT	NA	NA	NA	NA	NA	NA
IMPLICITTOP	NA	NA	NA	NA	NA	NA
TOPMANAGER	2	1.00	NA	1.00	1.00	1.00
DIVMANAGER	2	1.00	0.00	1.00	1.00	1.00
RSALES	2	0.93	0.00	0.92	0.93	0.93
SALESGROWTH	2	1.15	0.01	1.14	1.15	1.16
CATEGORY = 4						
TB	3	0.38	0.14	0.25	0.38	0.53
IMPLICITNEXT	NA	NA	NA	NA	NA	NA
IMPLICITTOP	NA	NA	NA	NA	NA	NA
TOPMANAGER	3	0.33	0.58	0.00	0.00	1.00
DIVMANAGER	3	1.00	0.00	1.00	1.00	1.00
RSALES	3	0.40	0.44	0.13	0.16	0.92
SALESGROWTH	3	1.27	0.23	1.12	1.15	1.53
CATEGORY = 5						
TB	22	0.38	0.10	0.34	0.38	0.38
IMPLICITNEXT	3	0.61	0.12	0.48	0.63	0.72
IMPLICITTOP	3	0.61	0.12	0.48	0.63	0.72
TOPMANAGER	23	0.48	0.51	0.00	0.00	1.00
DIVMANAGER	23	0.22	0.42	0.00	0.00	0.00
RSALES	23	0.41	0.40	0.10	0.24	0.94
SALESGROWTH	23	1.15	0.18	1.08	1.10	1.19
CATEGORY = 6						
TB	52	0.36	0.25	0.23	0.30	0.40
IMPLICITNEXT	9	0.83	1.14	0.24	0.29	1.47
IMPLICITTOP	9	0.94	1.08	0.40	0.46	1.55
TOPMANAGER	55	0.05	0.23	0.00	0.00	0.00
DIVMANAGER	55	0.11	0.31	0.00	0.00	0.00
RSALES	43	0.36	0.37	0.10	0.16	0.67
SALESGROWTH	43	1.13	0.14	1.07	1.12	1.17
CATEGORY = 7						
TB	85	0.31	0.23	0.17	0.25	0.40
IMPLICITNEXT	18	0.73	0.55	0.33	0.60	0.86
IMPLICITTOP	18	1.08	1.77	0.44	0.60	0.86

(continued on next page)

TABLE 4 (continued)

	n	Mean	Std.	Q1	Median	Q3
TOPMANAGER	85	0.00	0.00	0.00	0.00	0.00
DIVMANAGER	85	0.05	0.21	0.00	0.00	0.00
RSALES	74	0.20	0.28	0.03	0.10	0.21
SALESGROWTH	74	1.26	0.86	1.06	1.12	1.19
CATEGORY = 8						
TB	203	0.26	0.20	0.15	0.21	0.29
IMPLICITNEXT	54	0.43	0.18	0.30	0.45	0.56
IMPLICITTOP	54	0.71	0.23	0.53	0.64	0.93
TOPMANAGER	208	0.00	0.00	0.00	0.00	0.00
DIVMANAGER	208	0.04	0.20	0.00	0.00	0.00
RSALES	174	0.25	0.33	0.04	0.12	0.26
SALESGROWTH	174	1.14	0.18	1.06	1.11	1.17
CATEGORY = 9						
TB	323	0.20	0.12	0.11	0.16	0.23
IMPLICITNEXT	105	0.41	0.15	0.28	0.41	0.49
IMPLICITTOP	105	0.64	0.15	0.53	0.65	0.71
TOPMANAGER	331	0.00	0.00	0.00	0.00	0.00
DIVMANAGER	331	0.01	0.09	0.00	0.00	0.00
RSALES	279	0.19	0.26	0.04	0.10	0.20
SALESGROWTH	279	1.13	0.19	1.04	1.12	1.18
CATEGORY = 10						
TB	404	0.18	0.09	0.11	0.14	0.20
IMPLICITNEXT	179	0.38	0.08	0.33	0.37	0.41
IMPLICITTOP	179	0.58	0.09	0.55	0.58	0.62
TOPMANAGER	425	0.00	0.00	0.00	0.00	0.00
DIVMANAGER	425	0.00	0.07	0.00	0.00	0.00
RSALES	334	0.26	0.34	0.03	0.12	0.26
SALESGROWTH	334	1.16	0.72	1.07	1.12	1.19

[a] See Table 3 for variable definitions. The descriptive statistics are calculated for the pooled sample across countries, excluding 18 expatriate observations.

determines the implicit incentives. On the other hand, managers who are further down in the organization face lower compensation increases by moving up the initial levels, which are weighted most heavily in the computation of the option value.

The variable *TOPMANAGER* indicates whether a manager is among the highest-ranking managers in the respective country organization. Table 1 shows that the highest-level managers are concentrated in job-rating categories 2 through 6. Similarly, the descriptive statistics for *DIVMANAGER* indicate that the proportion of managers who are in charge of a local division is higher in the higher-level job-rating categories.

RSALES captures the relative importance of the division or subdivision with which a manager is affiliated, measured by the ratio of the unit's sales divided by the total sales of the respective country. As one would expect, the summary statistics for *RSALES* largely indicate that managers in lower job-rating categories are affiliated with smaller units. The median values for *RSALES* are

around 0.11 for categories 7 through 10, around 0.20 for categories 4 through 6, and over 0.90 for categories 2 and 3.

The median values for *SALESGROWTH*, which are measured by the ratio of budgeted sales for 2008 to actual sales for 2007 for the respective unit, largely indicate an upward trend as managers move to higher-level job categories. Managers in categories 2 through 4 are in charge of units that experience growth between 15 percent and 20 percent; in contrast, the expected unit growth for managers in categories 5 through 10 is around 11 percent. This finding is consistent with the notion that units that face higher growth opportunities are managed by higher-level managers who are delegated more decision-making authority.

IV. ANALYSES AND RESULTS

Strength of Implicit Incentives

The aim of this study is to analyze whether explicit incentives that are provided by variable-pay-based schemes are adjusted, based on the level of implicit incentives that are provided by the possibility of moving to a higher-level position in the hierarchy. The overall strategy that I use to address this question is to compare the explicit incentives of the local managers in the different countries. Given the organizational structure of the company, managers in the same job-rating category can reasonably be assumed to hold fairly comparable positions across countries. However, there is likely to be substantial variation in the strength of the implicit incentives that managers in different countries face due to the following reasons.

First, as can be observed from Table 1, the hierarchical structures differ across countries. Specifically, the countries vary with respect to the job category of the highest-ranking manager. In interviews with the author, the company contact emphasized that these differences are largely attributable to differences in the sizes of the units in the individual countries. Managers who have more organizational levels left "to climb" have stronger implicit incentives, all else equal.

Second, variation in the pay structures across job-rating categories and variation in the probabilities of getting promoted in the different countries can also result in differences in the implicit incentives that the managers face. Managers who face larger compensation increases upon being promoted and who have higher chances of getting promoted face stronger implicit incentives, *ceteris paribus*.

In order for the managers' implicit incentives to be determined by features of their local country organizations, their career paths inside this company have to be confined to their respective countries. This assumption seems reasonable for this setting for the following reason. The company pursues a strategy of being "multi-domestic," in the sense that the individual country organizations and their managers are expected to be very familiar with the local markets, governments, and infrastructures in order to be able to respond quickly to changes in the local environments. Part of this strategy is to employ local managers. Inspection of the data reveals that, aside from the 18 expatriates discussed above, the vast majority of managers are local. The company's use of predominantly domestic managers is consistent with other examples in the literature (Brickley et al. 2009). Thus, the individual country organizations appear to be fairly segregated, making it reasonable to assume that the majority of the managers in the sample do not move between countries.

Table 5 reports the correlation among the variables used in the analysis. *TB*, which captures the strength of the explicit incentives, is significantly positively correlated with the job-rating category of the highest-ranking manager in the respective country, but is not significantly correlated with either of the measures capturing the strength of the implicit incentives (*IMPLICITNEXT*, *IMPLICITTOP*). With respect to these preliminary findings, one should keep in mind that it is likely that incentive compensation practices systematically vary across countries for unobserved reasons. As expected, there is a significant correlation between *TB* and measures that

TABLE 5

Correlations among the Variables[a]

	TB	*HIGHEST-POSITION*	*IMPLICIT-NEXT*	*IMPLICIT-TOP*	*CATEGORY*	*Log(RSALES)*	*SALES-GROWTH*	*TOP-MANAGER*	DIV-MANAGER
TB									
HIGHESTPOSITION	0.46***								
IMPLICITNEXT	0.01	0.09*							
IMPLICITTOP	−0.03	0.09*	0.76***						
CATEGORY	−0.36***	−0.03	−0.32***	−0.22***					
Log(RSALES)	−0.03	−0.03	0.20***	0.07	−0.11***				
SALESGROWTH	−0.09**	−0.16***	0.04	0.03	−0.06*	0.04			
TOPMANAGER	0.23***	0.10***	NA	NA	−0.40***	0.16***	−0.01		
DIVMANAGER	0.19***	0.05	0.20***	0.27***	−0.32***	0.05	0.02	0.17***	

*, **, *** Significant at 10 percent, 5 percent, and 1 percent, respectively, all based on two-tailed tests.

[a] See Table 3 for variable definitions. The statistics are calculated for the pooled sample across countries, excluding 18 expatriate observations.

indicate that the manager has a high rank in the respective organization (*CATEGORY*, *TOPMANAGER*, *DIVMANAGER*).

The two measures capturing the strength of the implicit incentives, *IMPLICITNEXT* and *IMPLICITTOP*, are highly correlated with each other; they are also significantly correlated with measures that indicate that the manager has a high rank in the respective organization (*CATEGORY*, *DIVMANAGER*). The latter finding is consistent with the picture that emerges from the descriptive statistics in Table 3. Namely, the average values for the strength of the implicit incentives are increasing as one moves up the organizational ladder.

The hypothesis predicts that the explicit incentives provided by the company's compensation scheme are decreasing in the strength of the promotion-based implicit incentives that the manager faces. In order to investigate the hypothesis, I employ *TB*, which captures the ratio of the expected bonus to base salary, as the dependent variable. As discussed above, I employ several measures to capture the strength of the implicit incentives that the manager faces. Specifically, the different measures capturing the strength of the implicit incentives are the job-rating category of the highest-ranking manager in a respective country (*HIGHESTPOSITION*) and measures that take compensation differentials between adjacent job-rating categories as well as promotion possibilities into consideration (*IMPLICITNEXT*, *IMPLICITTOP*).

As mentioned above, it is plausible that incentive compensation practices vary systematically across countries for unobserved reasons. In order to control for such country-specific differences, I subtract from *TB* a measure that captures the median incentive intensity for mid-level managers for a large number of companies in the respective country. Specifically, the measure, which was provided by the company's consulting company, captures the median incentive intensity for managers who hold comparable positions in industrial corporations. The managers hold positions that are comparable to jobs that are ranked at level 10 in the company that is studied here. The consulting company computed the median incentive intensity using data for 30 to 150 companies in a given country. In order to control for systematic differences in the level of growth in the different countries, I adjust the variable *SALESGROWTH* for the respective country's growth in GDP from 2007 to 2008.

Since the sample includes measures that vary at the country level but that are constant for all managers within a country as well as measures that vary across managers, the structure of the data set can be described as hierarchical, with managers nested in countries considered lower in the hierarchy than countries (Bryk and Raudenbush 1992). Specifically, the job rating for the highest-ranking manager in a country, which is captured by the variable *HIGHESTPOSITION*, varies only at the country level.[12] Expressed in terms of the manager and country level of analysis, the model is as follows:[13]

Level 1:

$$\begin{aligned} TB_{ij} = {} & \beta_{0j} + \beta_1(IMPLICITNEXT_{ij} \text{ or } IMPLICITTOP_{ij}) + \beta_2 CATEGORY_{ij} \\ & + \beta_3 Log(RSALES_{ij}) + \beta_4 SALESGROWTH_{ij} + \beta_5 TOPMANAGER_{ij} \\ & + \beta_6 DIVMANAGER_{ij} + \varepsilon_{ij} \end{aligned}$$

for manager i in country j, where $\varepsilon_{ij} \sim N(0, \sigma^2)$.

Level 2:

$$\beta_{0j} = \gamma_0 + \gamma_1 HIGHESTPOSITION_j + u_{0j}$$

---

[12] See Anderson et al. (2000) for an application of hierarchical linear models in the accounting literature.
[13] Because the distribution of *RSALES* is skewed, I use a log-transformation.

for country j, where $u_{0j} \sim N(0, \tau^2)$.

Combined in expressed form, the model is as follows:

$$TB_{ij} = \gamma_0 + \gamma_1 HIGHESTPOSITION_j + \beta_1(IMPLICITNEXT_{ij}/IMPLICITTOP_{ij}) + \beta_2 CATEGORY_{ij} + \beta_3 Log(RSALES_{ij}) + \beta_4 SALESGROWTH_{ij} + \beta_5 TOPMANAGER_{ij} + \beta_6 DIVMANAGER_{ij} + u_{0j} + \varepsilon_{ij} \quad (1)$$

Table 6 shows the results of the estimation of Equation (1). I first estimate the model using the job-rating category of the highest-ranking manager in the respective country as the measure capturing the strength of the implicit incentives (Model I). The hypothesis predicts that the coefficient on *HIGHESTPOSITION*, β_1, is significantly positive. After controlling for the manager's job-rating category, the managers' implicit incentives are expected to be lower when the highest-ranking manager in the respective country has a lower job-rating category. The results indicate that, consistent with the prediction, the strength of the explicit incentives provided by the company's bonus plan is higher for managers who have fewer organizational levels left to climb.

In particular, the coefficient on *HIGHESTPOSITION* has a value of 0.025 and is significant at the 5 percent level, indicating that a manager's expected bonus decreases, on average, by approximately 2.5 percent of salary for each hierarchical level that s/he climbs. For example, a manager who has three organizational levels left to climb has an expected bonus that is lower by 2.5 percent of salary than the expected bonus of a manager holding a similar position but who has only two organizational levels left to climb. Given that the mean (median) values of the salary-scaled expected bonus are 0.36 (0.30) for managers who are at hierarchical level 6, the difference appears to be economically meaningful.

Although the findings with respect to the variable *HIGHESTPOSITION* are consistent with the hypothesis, they may be attributable to alternative explanations. Specifically, it is possible that the performance measures used in the bonus plan better reflect the actions of the managers who are closer to the top of their organization. Moreover, it is conceivable that managers who have fewer organizational levels left to climb have more decision-making authority that is not fully captured by the manager's job-rating category.

Model II is estimated by using *IMPLICITNEXT*, which is intended to capture the implicit incentives that the manager faces when focusing on the next-higher job-rating category. I predict a negative coefficient on *IMPLICITNEXT*. The explicit incentives that are provided by the company's bonus plan are expected to be lower when the compensation differential and/or the probability of getting promoted, which are captured by *IMPLICITNEXT*, are higher. The results obtained from estimating Model II are consistent with this prediction. Specifically, the coefficient on *IMPLICITNEXT* has a value of −0.079 and is significant at the 5 percent level. For example, if the median tenure is shorter by one year for managers at level 9, the value of *IMPLICITNEXT* increases by 0.29, which translates into an expected bonus that is lower by 2.2 percent of salary, on average. Similarly, a 50 percent increase in the median expected cash pay at the next level for managers in category 9 translates into an expected bonus that is lower by 1.8 percent of salary, on average.

Model III is estimated by using *IMPLICITTOP*, which is intended to capture the implicit incentives that the manager faces when considering the entire organizational ladder. The prediction and findings are consistent with the findings for Model II. In particular, the coefficient on *IMPLICITTOP* has a value of −0.047 and is significant at the 5 percent level. For example, if the median tenure is shorter by one year for managers at level 9, the value of *IMPLICITTOP* increases by 0.43, which translates into an expected bonus that is lower by 2.0 percent of salary, on average. Similarly, a 50 percent increase in the median expected cash pay at the next level for managers in

TABLE 6

Effect of the Strength of Implicit Incentives on Explicit Incentives[a]

$$TB_{ij} = \gamma_0 + \gamma_1 HIGHESTPOSITION_j + \beta_1(IMPLICITNEXT_{ij} \text{ or } IMPLICITTOP_{ij}) + \beta_2 CATEGORY_{ij} + \beta_3 Log(RSALES_{ij}) + \beta_4 SALESGROWTH_{ij} + \beta_5 TOPMANAGER_{ij} + \beta_6 DIVMANAGER_{ij} + u_{0j} + \varepsilon_{ij}$$

		Model I	Model II	Model III
Variable[b]		**Coefficient (z-statistic)**	**Coefficient (z-statistic)**	**Coefficient (z-statistic)**
Intercept		0.243 ***	0.605 ***	0.577 ***
		(4.40)	(2.93)	(2.75)
HIGHESTPOSITION	(+)	0.025 **		
		(2.10)		
IMPLICITNEXT	(−)		−0.079 **	
			(−2.45)	
IMPLICITTOP	(−)			−0.047 **
				(−2.11)
CATEGORY	(−)	−0.033 ***	−0.049 ***	−0.046 ***
		(−5.85)	(−4.42)	(−4.30)
Log(RSALES)	(+)	−0.001	−0.010	−0.012 *
		(−1.63)	(−1.42)	(−1.73)
SALESGROWTH	(+/−)	−0.019	−0.075	−0.076
		(−1.40)	(−1.22)	(−1.16)
TOPMANAGER	(+)	0.040 **	NA	NA
		(2.37)		
DIVMANAGER	(+)	0.026	0.172	0.183
		(1.52)	(1.25)	(1.26)
S.D. of Intercept (u_0)		0.071 ***	0.148 ***	0.140 ***
Number of observations		877	257	257
Pseudo R^2		39.85%	22.65%	23.73%

*, **, *** Significant at 10 percent, 5 percent, and 1 percent, respectively, all based on two-tailed tests.

[a] Expatriates are excluded from the regression models. Reported are the coefficients from the models with z-statistics in parentheses. The models are estimated using standard errors that are clustered by country.

[b] *TB* is calculated by subtracting the median incentive intensity in the respective country from *TB* as defined in Table 3; *SALESGROWTH* is calculated by subtracting the respective country's growth in GDP from *SALESGROWTH* as defined in Table 3; *Log(RSALES)* is natural logarithm of *RSALES*. The remaining variables are defined in Table 3.

category 9 translates into an expected bonus that is lower by 1.0 percent of salary, on average.

The findings with respect to the variables *IMPLICITNEXT* and *IMPLICITTOP* are consistent with the hypothesis. Nevertheless, it is possible that alternative explanations influence the findings. For example, it is possible that managers who face only small compensation increases upon promotion to the next level have relatively more decision-making authority.

With respect to the control variables in Table 6, the coefficient on the manager's job-rating category has a statistically significant value of around −0.03 to −0.04 across the three specifica-

tions. The findings indicate that a manager's expected bonus increases by 3–4 percent of salary, on average, when s/he moves to the next job-rating category. In Model I, the coefficient on the indicator variable, which indicates whether the manager is among the highest-ranking managers, is significantly positive, indicating that top-ranking managers receive higher explicit incentives. Specifically, the coefficient on *TOPMANAGER* has a value of 0.040, which suggests that top-ranking managers have expected bonus payments that are higher by 4 percent of salary, on average. The coefficients on the remaining control variables are insignificant at conventional levels. Specifically, the coefficient on *DIVMANAGER* is not significant at conventional levels. A potential explanation is that the effect is subsumed by the variables *TOPMANAGER* and *CATEGORY*.

In all three models, the standard deviation of the intercept is statistically significant. This finding indicates there is significant country-level variation around the intercept. Stated differently, there remains significant unexplained variation at the country-level.

It is conceivable that the results in Table 6 are driven by the number of organizational levels that the manager has left to climb. In other words, it is possible that the results in Models II and III with respect to the variables *IMPLICITNEXT* and *IMPLICITTOP* are driven by the manager's hierarchical level. In order to isolate the strength of the implicit incentives deriving from compensation increases and promotion probabilities, I re-estimate Models II and III from Table 6 after including the measure *HIGHESTPOSITION*, which controls for the number of organizational levels that the manager has left to climb. The results in Table 7 indicate that, even after controlling for the number of organizational levels that are left for the manager to climb, explicit incentives are stronger when promotion-based implicit incentives are weaker.

In summary, the results in Table 6 and Table 7 indicate that the explicit incentives provided by the company's bonus plan are higher when the manager has fewer organizational levels left to climb, when s/he faces lower implicit incentives from moving to the next organizational level, and when s/he faces lower implicit incentives from moving to the top of the organization. These findings support the hypothesis that explicit incentives provided by variable pay schemes are stronger when promotion-based implicit incentives are weaker. In a broader sense, the results are consistent with the notion that implicit promotion-based incentives are taken into consideration in designing explicit incentive contracts.

Robustness Tests

The inferences are robust to sensitivity tests. First, the results are robust to estimating an OLS regression on the pooled sample where the intercept is treated as non-random. Second, I repeat the analyses by including an indicator variable for each of the five different divisions. The results are virtually identical. Third, in order to control for the manager's career horizon (Gibbons and Murphy 1992), I include the manager's age as an independent variable in the regression analyses, which does not change the results. I also repeat the analyses using the ratio of the actual bonus that was paid out to base salary as the dependent variable in order to control for potential differences in the countries' bonus payout practices. The inferences remain unchanged. Moreover, the inferences are robust to including the unit's sales as a measure of the unit's size in the regression models. The results also remain unchanged when the variable capturing the manager's job-rating category, *CATEGORY*, is replaced by indicator variables in order to address potential nonlinearities.

V. CONCLUSION

This study re-examines the hypothesis that the explicit, compensation-based incentives of mid-level managers are adjusted to the level of implicit, promotion-based incentives. Specifically,

TABLE 7

Effect of the Strength of Implicit Incentives on Explicit Incentives[a]

$$TB_{ij} = \gamma_0 + \gamma_1 HIGHESTPOSITION_j + \beta_1(IMPLICITNEXT_{ij} \text{ or } IMPLICITTOP_{ij}) + \beta_2 CATEGORY_{ij} + \beta_3 Log(RSALES_{ij}) + \beta_4 SALESGROWTH_{ij} + \beta_5 TOPMANAGER_{ij} + \beta_6 DIVMANAGER_{ij} + u_{0j} + \varepsilon_{ij}$$

Variable[b]		I Coefficient (z-statistic)	II Coefficient (z-statistic)
Intercept		0.420 ***	0.402 **
		(4.51)	(3.56)
HIGHESTPOSITION	(+)	0.031	0.031
		(0.95)	(0.98)
IMPLICITNEXT	(−)	−0.086 ***	
		(−4.30)	
IMPLICITTOP	(−)		−0.057 **
			(−2.18)
CATEGORY	(−)	−0.046 ***	−0.045 ***
		(−5.01)	(−4.62)
Log(RSALES)	(+)	−0.003	−0.005
		(−0.27)	(−0.48)
SALESGROWTH	(+/−)	−0.060	−0.065
		(−0.92)	(−0.93)
TOPMANAGER	(+)	NA	NA
DIVMANAGER	(+)	0.190	0.204
		(1.28)	(1.31)
S.D. of Intercept (u_0)		0.130 ***	0.130 ***
Number of observations		257	257
Pseudo R^2		27.04%	28.27%

*, **, *** Significant at 10 percent, 5 percent, and 1 percent, respectively, all based on two-tailed tests.

[a] Expatriates are excluded from the regression models. Reported are the coefficients from the models with z-statistics in parentheses. The models are estimated using standard errors that are clustered by country.

[b] *TB* is calculated by subtracting the median incentive intensity in the respective country from *TB* as defined in Table 3; *SALESGROWTH* is calculated by subtracting the respective country's growth in GDP from *SALESGROWTH* as defined in Table 3; *Log(RSALES)* is natural logarithm of *RSALES*. The remaining variables are defined in Table 3.

this study revisits the theoretical argument that explicit incentives are optimally stronger in situations that pose weaker implicit, promotion-based incentives (Gibbons and Murphy 1992; Gibbs 1995).

The analyses in this study are based on compensation data from a large multinational corporation. This setting provides an opportunity to observe variation in the strength of implicit incentives because the sample is comprised of managers with comparable jobs but who face varying levels of implicit incentives since they are positioned at different hierarchical levels in their respective organization, their promotion possibilities vary, and they experience different levels of compensation increases upon promotion.

Regression analyses show that the incentives provided by the company's bonus plan are stronger for managers who are positioned at higher hierarchical levels, who face weaker implicit incentives from getting promoted to the next level, and who face weaker implicit incentives from getting promoted to the top of the organization, after controlling for the position's scope and level of accountability. These findings are consistent with the notion that implicit promotion-based incentives are taken into consideration in designing explicit incentive contracts, as proposed in the theoretical literature. More precisely, the evidence supports the hypothesis that explicit incentives are optimally stronger in situations with weaker implicit incentives (Gibbons and Murphy 1992; Gibbs 1995).

REFERENCES

Anderson, S., D. Glenn, and K. Sedatole. 2000. Sourcing parts of complex products: Evidence on transactions costs, high-powered incentives and ex-post opportunism. *Accounting, Organizations and Society* 25 (8): 723–749.

Baiman, S., D. Larcker, and M. Rajan. 1995. Organizational design for business units. *Journal of Accounting Research* 33 (Autumn): 593–616.

Baker, G., M. Gibbs, and B. Holmstrom. 1994. The internal economics of the firm: Evidence from personnel data. *The Quarterly Journal of Economics* 109 (4): 881–919.

———, and B. Hall. 2004. CEO incentives and firm size. *Journal of Labor Economics* 22 (October): 767–798.

Brickley, J., C. Smith, and J. Zimmerman. 2009. *Managerial Economics and Organizational Architecture*. New York, NY: McGraw-Hill Companies.

Bryk, A. S., and S. W. Raudenbush. 1992. *Hierarchical Linear Models: Applications and Data Analysis*. Newbury Park, CA: Sage Publications.

Bushman, R., and A. Smith. 2001. Financial accounting information and corporate governance. *Journal of Accounting and Economics* 32 (1–3): 237–333.

Campbell, D. 2008. Nonfinancial performance measures and promotion-based incentives. *Journal of Accounting Research* 46 (2): 297–332.

Ederhof, M. 2010. Discretion in bonus plans. *The Accounting Review* 85 (6): 1921–1949.

Gaver, J., and K. Gaver. 1993. Additional evidence on the association between the investment opportunity set and corporate financing, dividend, and compensation policies. *Journal of Accounting and Economics* 16 (1–3): 125–160.

Gibbons, R., and K. Murphy. 1992. Optimal incentive contracts in the presence of career concerns: Theory and evidence. *The Journal of Political Economy* 100 (June): 468–505.

Gibbs, M. 1995. Incentive compensation in a corporate hierarchy. *Journal of Accounting and Economics* 19 (2–3): 247–277.

——— 1996. Promotions and incentives. Working paper, University of Chicago.

Holmstrom, B. 1979. Moral hazard and observability. *The Bell Journal of Economics* 10 (Spring): 74–91.

Holthausen, R., D. Larcker, and R. Sloan. 1995. Business unit innovation and the structure of executive compensation. *Journal of Accounting and Economics* 19 (2–3): 279–313.

Indjejikian, R., and D. Nanda. 2002. Executive target bonuses and what they imply about performance standards. *The Accounting Review* 77 (4): 793–819.

Jensen, M., and W. Meckling. 1976. Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics* 3 (4): 305–360.

Kahn, L., and P. Sherer. 1990. Contingent pay and managerial performance. *Industrial & Labor Relations Review* 43 (February): 107S–120S.

Lambert, R., and D. Larcker. 1987. An analysis of the use of accounting and market measures of performance in executive compensation contracts. *Journal of Accounting Research* 25 (Supplement): 85–125.

———, ———, and K. Weigelt. 1993. The structure of organizational incentives. *Administrative Science Quarterly* 38 (September): 438–461.

Lazear, E., and S. Rosen. 1981. Rank order tournaments as optimum labor contracts. *The Journal of Political Economy* 89 (October): 841–864.

Liang, P., M. Rajan, and K. Ray. 2008. Optimal team size and monitoring in organizations. *The Accounting Review* 83 (3): 789–822.

Nagar, V. 2002. Delegation and incentive compensation. *The Accounting Review* 77 (2): 379–395.

Ortin-Angel, P., and V. Salas-Fumas. 1998. Agency-theory and internal-labor-market explanations of bonus payments: Empirical evidence from Spanish firms. *Journal of Economics & Management Strategy* 7 (Winter): 573–613.

Prendergast, C. 2000. What trade-off of risk and incentives? *The American Economic Review* 90 (2): 421–425.

——— 2002. The tenuous trade-off between risk and incentives. *The Journal of Political Economy* 110 (October): 1071–1102.

Rosen, S. 1986. Prizes and incentives in elimination tournaments. *The American Economic Review* 76 (September): 701–715.

Smith, C., and R. Watts. 1992. The investment opportunity set and corporate financing, dividend, and compensation policies. *Journal of Financial Economics* 32 (3): 263–292.

Wulf, J. 2007. Authority, risk, and performance incentives: Evidence from division manager positions inside firms. *The Journal of Industrial Economics* 55 (March): 169–196.

THE ACCOUNTING REVIEW
Vol. 86, No. 1
2011
pp. 155–184

American Accounting Association
DOI: 10.2308/accr.00000010

The Role of Organizational Absorptive Capacity in Strategic Use of Business Intelligence to Support Integrated Management Control Systems

Mohamed Z. Elbashir
The Australian National University

Philip A. Collier
The University of Melbourne

Steve G. Sutton
University of Central Florida

ABSTRACT: This study examines the influence of organizational controls related to knowledge management and resource development on assimilation (i.e., strategic integration and use) of business intelligence (BI) systems. BI systems use analytics and performance management concepts to leverage enterprise system databases and provide core management control system (MCS) capability. Our results indicate that organizational absorptive capacity (i.e., the ability to gather, absorb, and strategically leverage new external information) is critical to establishing appropriate technology infrastructure and to assimilating BI systems for organizational benefit. Further, findings show that while top management plays a significant role in effective deployment of BI systems, their impact is indirect and a function of operational managers' absorptive capacity. In particular, this indirect effect suggests that leveraging BI systems is driven from the bottom up as opposed to the top down. This differentiates BI from other isolated strategic MCS innovations that have traditionally been viewed as top-management driven.

Keywords: *business intelligence; management control systems; knowledge creation; absorptive capacity; management accounting systems; business analytics; corporate performance management; enterprise systems; enterprise resource planning systems.*

The authors thank Stewart Leech, Michael Davern, David Smith, Vicky Arnold, Shannon Anderson, Carlin Dowling, Uday Murthy, Anol Bhattacherjee, Andrea Kelton, Rod Dilnutt, Robyn King, and workshop participants at the Australian National University, University of Melbourne, University of Queensland, University of South Florida, and at the 2008 European Conference on Accounting Information Systems and the 2009 Accounting & Finance Association of Australia and New Zealand Annual Meeting for their valuable feedback at various stages of the research.

Editor's note: Accepted by James Hunton.

Submitted: April 2008
Accepted: June 2010
Published Online: January 2011

Data Availability: *Data are available through the first author, subject to limitations set by the Faculty Human Ethics Advisory Group and the vendor providing customer information.*

I. INTRODUCTION

Enterprise systems (i.e., integrated systems such as enterprise resource planning [ERP]) are fundamentally tied to the work of accounting. Enterprise systems have transformative implications for the integration of organizational information and control systems governing organizations (Chapman 2005; Chapman and Kihn 2009). Unfortunately, research indicates that most organizations' management control systems (MCS[1]) have not leveraged this potential (e.g., Granlund and Malmi 2002; Dechow and Mouritsen 2005; Quattrone and Hopper 2005; Rom and Rohde 2007), but have simply used these systems to support existing MCS (Chapman and Kihn 2009). Enterprise systems are simply a resource that is made available and must be effectively leveraged for MCS enhancements to actually accrue (Chapman and Kihn 2009).

Business intelligence (BI) systems are widely viewed as the innovation that can leverage the wealth of data encapsulated in an enterprise system and support the anticipated transformation to a broader and more detailed MCS (Brignall and Ballantine 2004; Williams 2004; Carte et al. 2005; Kay 2006; Gnatovich 2007; Robertson et al. 2007). BI systems provide business analytics and corporate performance management reporting capabilities, the fundamental information for an effective MCS in a technology-driven business environment.[2] Davila and Foster (2007) identify 46 specific categories of MCS. The extensive set of pre-built reports and metrics included in most BI systems provide support for all of these categories.[3] The purpose of this study is to examine the deployment of BI systems (an emerging MCS innovation) in order to better understand how an organization's controls related to knowledge management and resource development can lead to higher levels of BI assimilation.

Prior research suggests that integration of the enterprise system and MCS is heavily dependent on an organization's people, such that both the social and technical aspects of organizations should be considered in future research (Chapman and Kihn 2009). This view is consistent with Chenhall (2005), who argues that future research on MCS adoption needs to specifically consider how cultural controls influence implementation of an MCS innovation. Simon (1995, 34) defines value (i.e., cultural) controls as the "explicit set of organizational definitions that senior managers communicate formally and reinforce systematically to provide basic values, purpose, and direc-

---

1 We refer to MCS as the formal, information-based routines and procedures that provide managers with measures, performance indicators, and procedures to maintain or alter patterns in the organizational activities to ensure that they are consistent with the organizational objectives and strategies (Simon 1995; Malmi and Brown 2008).

2 Business Analytics provide the functionality to extract enterprise data and create a broad range of balanced scorecards, key performance indicators, total quality management metrics, and activity-based costing. Primarily they are associated with cybernetic controls while also providing some support for compensation/reward controls and administrative controls. Corporate Performance Management (CPM) capabilities provide support for budgeting, financial planning, consolidation, and enterprise planning. CPM also provides predefined reports for ABC, scorecards, and budgets. Thus, CPM primarily extends across planning and cybernetic controls (Howard 2003; Williams and Williams 2007).

3 The specific BI system used by the organizations that we examine includes three broad applications: financial, supply chain, and customer analyses. Financial analysis focuses on accounts receivable (customer credit, performance, cash inflow, and organizational effectiveness), general ledger (financial ratio reporting and analysis, financial performance, organizational performance, and budget analysis), and accounts payable (cash outflow, performance, vendor account analysis, and organizational effectiveness). Supply chain analysis focuses on inventory (stock overview and valuation, inventory forecasting, inventory demand, material movement activity, physical inventory, material reservations, and organizational effectiveness), procurement (material expenditure, procurement vendor analysis, material demand, process effectiveness, and organizational effectiveness) and the accounts payable analysis shared with the financial analysis application. Customer analysis includes sales (functional performance, customer sales, product sales, channel sales, distribution functional performance, sales organizational effectiveness, and distribution organizational effectiveness) and accounts receivable that is shared with financial analysis (Howard 2003).

tion." Using that definition, we focus on one specific aspect of cultural control—that of control over knowledge management and resource development. The resulting culture determines the identification, absorption, and strategic application of new external information (hereafter, the level of *absorptive capacity*). Our focus includes the influence of the *top management team* (TMT) on the culture present at the operational manager level. By studying the level of absorptive capacity, we shed light on the interrelationship between cultural controls and deployment of an integrated MCS—i.e., MCS as a package (Malmi and Brown 2008).

The study of MCS as a package addresses another concern in the MCS innovation literature—the tendency to study MCS innovations in isolation. Chenhall (2003, 131) notes that MCS innovation studies are generally done in isolation. If specific accounting controls are systematically linked with other organizational controls, then research ignoring these connections may report spurious findings. BI systems provide a unique type of MCS innovation, in that they are not designed to support a single aspect of control (e.g., activity-based costing or balanced scorecards). Rather, BI systems provide powerful extraction capability to supplement substantial expansion of both planning and cybernetic controls, while also providing support for administrative and reward/compensation controls.[4] In our study, we focus on organizations' deployment of one BI system that is specifically designed to enhance MCS capability. At implementation, the system provides pre-developed metrics that can be configured to connect to the underlying databases of most leading ERP vendors. These metrics provide access to over 200 different pre-built reports using more than 500 key performance indicators and analytics answering over 2,900 business critical questions. The metrics include a broad array of performance measures for sales analysis, financial analysis, inventory and procurement analysis, and supply chain analysis—including a multitude of scorecard analyses (Howard 2003). Thus, amid concerns that MCS innovations have been studied in isolation from the overall package of MCS systems present in organizations (Chenhall 2005; Davila and Foster 2007; Malmi and Brown 2008), BI systems allow us to examine deployment of an integrated MCS.

We also address two general limitations that have been raised in regards to MCS adoption studies. First is the concern that studies have examined whether firms have adopted MCS, but those studies have not explored the variation in quality or depth of the use of MCS following that adoption (Davila and Foster 2007). Our research extends beyond adoption to assimilation, which includes the scope, use, and strategic integration of a system (Chatterjee et al. 2002). Second is the difficulty in moving beyond innovative case-based studies of MCS that can have limited generality (Davila and Foster 2005). Garnering access to multiple implementers of an innovation is generally quite difficult. Even when access may be possible, the variation in implementations often limits comparability. We avoided these problems by approaching one of the major BI vendors and securing their full client list. As a result, we acquire responses from a diverse sample of organizations implementing a common BI system.

[4] Chenhall (2003) differentiates MCS from management accounting systems (MAS) by noting the latter is a subcomponent of the former. Chenhall (2003) defines the MAS as the systematic use of management accounting practices such as budgeting or product costing. MCS is broader and encompasses MAS as well as cultural and administrative controls. When we refer to core MCS capability, MAS capability is the primary interest that is encompassed primarily by cybernetic and planning controls (Malmi and Brown 2008). Malmi and Brown (2008) define the various categories of control as follows: *culture controls* establish the values, beliefs, and social norms that influence employees' behavior; *cybernetic controls* provide quantitative measures for activities and processes, set performance standards, provide feedback processes, assess variance between goals and accomplishments, and influence individual behavior; *planning controls* set functional area goals, establish standards for assessing accomplishment, and assure alignment of goals across the functional areas of the organization; *reward/compensation controls* motivate and increase the performance of individuals; and *administrative controls* establish organizational structures, lines of accountability, and establish policies and procedures.

This study examines the role of top-management's and operational-level-management's absorptive capacity in facilitating assimilation of BI into the overall MCS. We develop an integrative model that theorizes relationships between the strategic and operational levels of absorptive capacity, and the impact on the existing IT infrastructure in order to improve our understanding of how organizations assimilate BI. We test the model using data collected from 347 business units that implemented the BI software. Our results indicate that organizational absorptive capacity is fundamental to both the readiness of the technology infrastructure for supporting BI integration and the successful assimilation into the overall MCS.

This study offers several contributions to the MCS literature. First, we examine the influence of the overall MCS on the adoption and strategic integration of BI, an MCS innovation. This is consistent with calls that advocate that MCS should be considered as a package, allowing MCS innovations to be examined in their context rather than in isolation. Second, in examining the cultural control aspect, we provide clarity to the role of TMT in assimilation of MCS innovation. By incorporating absorptive capacity, the result of cultural controls over knowledge management, we are able to show the critical intermediary role of operational management's absorptive capacity on assimilation of an MCS innovation. Third, our study moves beyond adoption of an MCS to consider the depth of strategic integration. By examining the assimilation of BI systems, we attain a better understanding of the strategic benefit that is derived from an MCS innovation. Fourth, we are able to establish that these relationships are consistent across a diverse range of organizations by examining one common BI system.

This study also contributes to the accounting information systems literature. First, for 30 years researchers have been focused on the normative development of integrated databases that unify enterprise-wide business event data (McCarthy 1982). This study provides empirical evidence of the benefit such integrated databases have for effective MCS. Second, our study provides insights into the strategic benefit that can accrue from standardized software extracting data from enterprise databases and providing both pre-specified and dynamic reporting capability. Third, we expand the general IT literature on the relationship between absorptive capacity and IT assimilation by focusing not only on the static knowledge component of absorptive capacity, but also the knowledge creation component. Our results indicate this latter component is critical to the assimilation of complex strategic systems such as BI.

Section II draws from theory on systems assimilation and the dynamic perspective of knowledge creation to develop a model of BI assimilation. Section III details the research methods, operationalization of constructs, and data collection process. Section IV details the results, and Section V discusses the implications.

II. THEORY AND HYPOTHESES

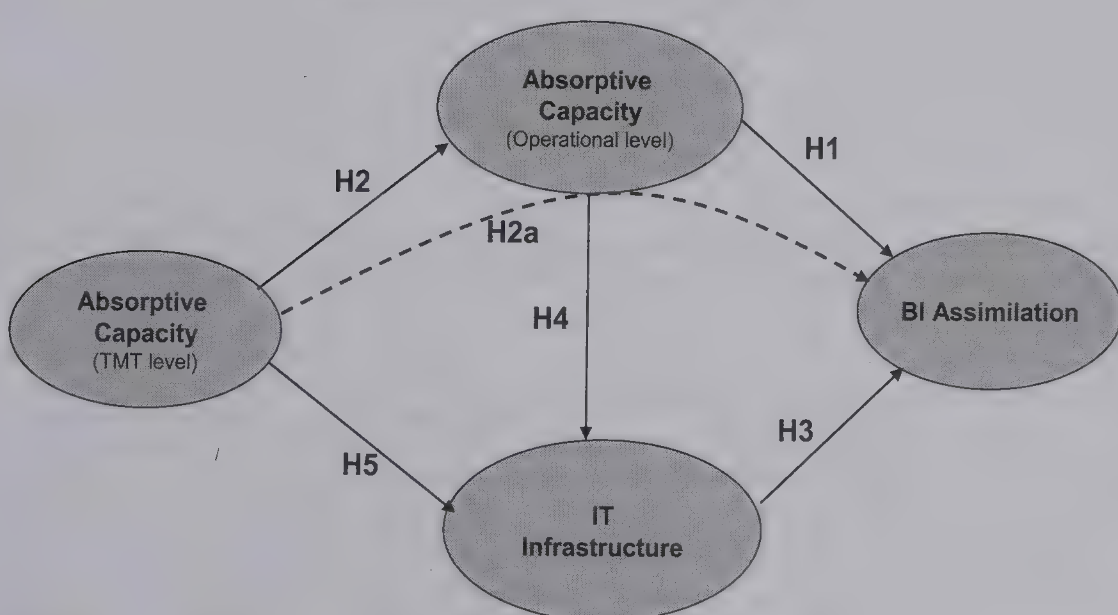
The adoption of enterprise systems and the resulting impact on MCS have been studied with great anticipation, given the potential for real benefits from integrated enterprise-wide data. Enterprise systems are expected to automate mundane MCS tasks and provide opportunity for broader-based MCS. These extended MCS are expected to enhance both management's strategic analyses and operational-level analyses as well as improve decision making (Sutton 2000; Granlund and Malmi 2002; Chapman 2005; Dechow and Mouritsen 2005; Quattrone and Hopper 2005; Arnold 2006). While this transformation is prevalent to some limited degree in some organizations (e.g., Granlund and Malmi 2002; Caglio 2003; Scapens and Jazayeri 2003; Quattrone and Hopper 2005), the general consensus is that little evolution in MCS has occurred. Rather, enterprise-wide data are often simply extracted for use by the same MCS modules that existed prior to implementation (Granlund and Malmi 2002; Dechow and Mouritsen 2005; Rikhardsson and Kraemmergaard 2006; Rom and Rohde 2007). Simply deploying enterprise systems appears insufficient to achieve significant enhancement of organizations' MCS (Chapman

and Kihn 2009). The professional literature similarly espouses the inadequacy of enterprise systems for facilitating the reporting and analyses necessary to support advanced MCS (e.g., Williams 2004; Kay 2006; Gnatovich 2007; Robertson et al. 2007; Williams 2008).

Enterprise systems provide highly integrative databases, but the ability for the average user to extract relevant data without the aid of specialized applications is limited. BI systems are designed to facilitate users in conducting detailed analyses of data contained in enterprise databases (Brignall and Ballantine 2004; Carte et al. 2005; Robertson et al. 2007; Williams 2008). BI systems hook into the underlying databases created by enterprise systems and provide a broad array of pre-specified reports and business analytics that form the information infrastructure necessary to support enhanced MCS capability.

By definition, assimilation studies are post-adoption studies that assume the decision to adopt a technology has been made, and acceptance and diffusion of the system is complete (Chatterjee et al. 2002). At an assimilation level, the interest is in whether a strategic technology has been integrated and used in a manner that provides strategic benefit, which is the focus of this study. Figure 1 provides an overview of our conceptual model. The model focuses on three critical organizational factors theorized to affect BI assimilation: TMT's absorptive capacity, operational managers' absorptive capacity, and sophistication of IT infrastructure. An organization's absorptive capacity is represented by its ability to recognize the value of new, external information, absorb it, and apply it for commercial ends (Cohen and Levinthal 1990). This focus on the absorptive capacity (Cohen and Levinthal 1990; Zahra and George 2002) of an organization's top-level and operational management is consistent with the views put forth by Chapman and Kihn (2009). This focus entails inclusion of the critical organizational components necessary for successfully leveraging enterprise systems. The joint effects of TMT and operational managers' absorptive capacity capture the effectiveness of cultural controls related to knowledge management.

FIGURE 1
Conceptual Model for BI Assimilation



Cultural Controls and the Development of Absorptive Capacity

Cultural controls relate to setting and instilling a set of values, beliefs, and social norms that are shared by an organization's members (Malmi and Brown 2008). Cultural controls encapsulate what Simons (1995, 34) identifies as "value" controls, and that he defines as "the explicit set of organizational definitions that senior managers communicate formally and reinforce systematically to provide basic values, purpose, and direction for the organization." Simons (1995, 34) notes further that these organizational definitions specify the "values and directions that senior managers want subordinates to adopt." Cultural controls play an important role in a broader view of MCS in that they set the boundaries and direction for core MCS capability (e.g., planning, cybernetic, and reward/compensation controls) and influence the role MCS plays in the organization (Chenhall 2003; Malmi and Brown 2008).

TMT represent a small group of the most influential executives with overall responsibility for an organization (Hambrick and Mason 1984). While MCS studies often highlight the critical role of TMT (Anderson and Young 1999; Bhimani 2003; Davila and Foster 2005; Naranjo-Gill and Hartman 2006, 2007; Davila and Foster 2007), TMT's knowledge has received limited attention in the MCS literature. TMT's knowledge is considered a primary indicator of TMT's competence (Armstrong and Sambamurthy 1999; Bassellier et al. 2003), yet research has failed to link TMT knowledge and assimilation (e.g., Armstrong and Sambamurthy 1999). We explore this phenomenon with the belief that prior studies, which have focused on a *static view*[5] of TMT's knowledge, fail to capture the knowledge-creation activities that should drive technology innovation.

TMT knowledge from a *dynamic view*[6] of organizational knowledge should provide a much better encapsulation of TMT's knowledge capabilities. In turn, a dynamic view should better capture the underlying theoretical reasons for the link between TMT and BI assimilation. An absorptive-capacity view focuses on the synergies of both TMT's knowledge and TMT's ability to put that knowledge into practice. From a theory perspective, TMT's absorptive capacity should be the key determinant of TMT's ability to provide effective leadership and support increased absorptive capacity at all levels of the organization. TMT's absorptive capacity is a key element for cultural controls to be effective in promoting organizational absorptive capacity.

Within an MCS context, operational managers can play a key role in driving the use of BI systems to support an organization's MCS needs. The effect of TMT's absorptive capacity on assimilation is therefore best conceptualized as indirect, mediated by operational management's absorptive capacity. Case-based evidence of the phenomenon can be interpreted from Caglio (2003), where the CFO became the driving force behind the design and implementation of the enterprise system. In the course of taking this lead, the CFO empowered the management accountants to be retrained to take on more of an analytic role for the TMT and a consulting role to a broad range of functional areas. Hence, the CFO's quick adaptation to the enterprise system and his rapid acquisition of relevant knowledge and capabilities allowed him to be an effective force in the implementation. However, the actual individuals implementing the changes that facilitated

---

[5] The *static view* of organizational knowledge perceives organizations as a stock of knowledge that an organization possesses (Nonaka 1994; Nonaka and Takeuchi 1995; Grant 1996; Cook and Brown 1999). Most assimilation studies that have drawn on the knowledge-based view follow the static view and capture the amount of knowledge that people possess. Cook and Brown (1999), among many scholars, criticize research on organizational knowledge which follows the static view (Nonaka 1994; Kim 1998; Nonaka and Toyama 2003). Cook and Brown (1999) refer to this traditional understanding of organizational knowledge as the *epistemology of possession* because organizational knowledge is treated as something people possess.

[6] The *dynamic perspective* of organizational knowledge is consistent with a large body of current knowledge management literature (Ditillo 2004; Vera-Munoz et al. 2006) that suggests understanding the organization's knowledge through its dynamic capabilities of knowledge creation rather than the stock of knowledge (static view) that organizations possess (Nonaka 1994; Grant 1996; Cook and Brown, 1999).

the transformation are the operational-level accountants acquiring the new knowledge to facilitate the transformation. The CFO was able to communicate his vision, providing a strong cultural control that helped transform behavior. This organization's experience is consistent with theorizations that organizations implementing controls promoting organizational absorptive capacity are better able to overcome knowledge barriers and effectively utilize new systems (Attewell 1992; Fichman and Kemerer 1997; Ravichandran 2005).

Cohen and Levinthal (1990), incorporating Attewell's (1992) views on organizational learning and innovation, posit that an organization's absorptive capacity facilitates learning and, in turn, drives innovation. As noted earlier, an organization's absorptive capacity is represented by its ability to recognize the value of new, external information, absorb it, and apply it for commercial ends (Cohen and Levinthal 1990).[7] Cohen and Levinthal (1990) suggest that *prior relevant knowledge* and *intensity of effort* are critical elements for developing effective absorptive capacity. Intensity of effort represents the energy that members of an organization commit to solving problems and creating new knowledge.[8]

Absorptive Capacity and BI Assimilation

Successfully assimilating a BI system that automates MCS capabilities aligned with business strategies is dependent upon the development of relevant knowledge and skills (Fichman and Kemerer 1997; Armstrong and Sambamurthy 1999). The development of relevant knowledge and skills within an organization is highly dependent on the cultural controls management has put in place and supported (Malmi and Brown 2008). These cultural controls are used by organizations to integrate individuals' absorptive capabilities into the organization's routine and practice (Nonaka 1994; Kim 1998; Van den Bosch et al. 1999). Organizational absorptive capacity reflects the competence attributed to an organization's members (Szulanski 1996; Zahra and George 2002; Bassellier et al. 2003).

Two levels of organizational absorptive capacity are critical to BI assimilation. *TMT's absorptive capacity* represents the collective ability of TMT members to recognize the value of new information gathered from both internal and external sources, absorb it, and apply it to support their leadership role in strategic planning and control. TMT's absorptive capacity is determined by the broader knowledge and expertise that its members possess as well as knowledge-creation activities, including interaction with external constituents, such as competitors, customers, and peers (Nambisan et al. 1999; Daghfous 2004). *Operational-level managers' absorptive capacity* refers to the ability of managers at the operational level to value new information, absorb it, and apply it to support the organization's business strategy and value chain activities. Absorptive capacity at the operational level is heavily influenced by cultural controls, including TMT inter-

---

7 This definition of absorptive capacity suggests that the organization's absorptive capacity is built upon three capabilities: value, absorption, and application (Cohen and Levinthal 1990; Zahra and George 2002). (1) *Value capability* refers to the organization's ability to recognize new information, whether received from internal or external sources. This capability requires the organization to possess prior relevant knowledge and expertise that will help to value/assess new information. (2) *Absorption capability* refers to the organization's ability to analyze, process, interpret, and understand the new information. (3) *Application capability* refers to the ability of the organization to use the new information to support an organization's activities and strategies. The three capabilities are combinative in nature as they build upon each other to create the dynamic capability of the organization. This also suggests that an organization's absorptive capacity is path-dependent, as prior relevant knowledge must exist in order to facilitate absorption and use of new information.

8 Kim (1998) suggests a matrix that describes the interaction between prior relevant knowledge and intensity of effort and the effect of that interaction on the organizational absorptive capacity. According to Kim (1998) the absorptive capacity of the organization will be at its highest level when both the intensity of effort (i.e., action) and prior relevant knowledge are high. The absorptive capacity will be at the lowest level when both prior relevant knowledge and knowledge-related effort are low. In this study we use the underlying concept articulated by this matrix to operationalize the absorptive capacity construct tested in our research model.

vention and focused knowledge-creation activities (Boynton et al. 1994; Fichman and Kemerer 1997; Jansen et al. 2005).

BI assimilation is largely focused on the support and integration of core MCS control processes (e.g., planning and cybernetic) that are of primary concern at the operational or line levels. These core control processes are generally thought of as the management accounting system (Chenhall 2003; Malmi and Brown 2008). We see parallels in the few reported successful enterprise system integration studies. For instance, Caglio (2003) notes the empowerment of management accountants across operational and line functions with their ability to leverage enterprise-wide data for a broad range of functional areas. We posit that operational-level absorptive capacity is a critical precursor to BI assimilation, leading to H1:

H1: Operational-level managers' absorptive capacity will positively enhance organizations' BI assimilation.

As noted earlier, TMT plays a key role in establishing the cultural controls that motivate and enable the development of absorptive capacity. TMT leadership roles can be viewed from two perspectives: external and internal (Ulrich and Wiersema 1989; Kakabadse et al. 1995). External leadership roles include the ability to interact with the changing environment and interpret this into internal vision (Daft and Weick 1984; Hambrick 1995). Internal leadership roles include the design and management of employees' actions that enable realization of an organization's vision (Kakabadse et al. 1995; Anderson and Young 1999; Caglio 2003; Chenhall and Euske 2007). This includes provision of knowledge-creation mechanisms that address knowledge gaps at both the TMT and operational levels and are necessary to support new strategies and enabling technologies (Keen 1991; Nonaka et al. 1998; Caglio 2003).

Identifying and remediating knowledge gaps is critical, as BI assimilation is dependent on operational-level managers understanding the full potential of BI systems. This requires operational-level managers to raise their IT literacy to a level conducive with effective deployment (Rikhardsson and Kraemmergaard 2006). This dependency on operational-level learning suggests the relation between TMT's absorptive capacity and BI assimilation is mediated by operational-level absorptive capacity. The second hypothesis is stated as:

H2: TMT's absorptive capacity will positively enhance operational level-managers' absorptive capacity.

H2a: TMT's absorptive capacity will positively enhance organizations' BI assimilation, through the operational-level managers' absorptive capacity.

IT Infrastructure Sophistication and BI Assimilation

IT infrastructure sophistication refers to "the extent to which an organization has diffused the key information technologies into its foundation for supporting business applications" (Armstrong and Sambamurthy 1999, 309). IT infrastructure sophistication reflects the diversity and integration of IT components necessary to support BI systems. BI is designed to leverage complex business data (Quattrone and Hopper 2005; Dechow and Mouritsen 2005) that are integrated with other business information in the creation of enterprise-wide databases (Granlund and Malmi 2002). Accordingly, we posit that organizations with the sophisticated IT infrastructure to support information systems integration and enterprise-wide data integration will be better able to successfully assimilate BI systems. This reasoning leads to H3:

H3: The sophistication of IT infrastructure will positively enhance an organization's BI assimilation.

Absorptive Capacity and IT Infrastructure Sophistication

IT users at the operational level are important sources of IT innovation (von Hippel 1994; Nambisan et al. 1999). In an enterprise systems environment, operational-level managers are capable of driving systems configuration (Caglio 2003; Rikhardsson and Kraemmergaard 2006; Byrne and Pierce 2007; Elbashir and Williams 2007). The manner in which IT infrastructure is deployed can directly affect the information available to support MCS objectives (Quattrone and Hopper 2005). Supporting broad-based MCS capabilities necessitates sophisticated infrastructure configurations only possible through high levels of operational-level absorptive capacity (Dechow and Mouritsen 2005). This reasoning leads to H4:

H4: Operational-level managers' absorptive capacity will positively enhance the IT infrastructure sophistication of the organization.

TMT contributes to various IT infrastructure-related activities including project planning, resource allocation, and user problem solving. Building sophisticated IT infrastructure to support various applications requires a sound understanding of the organization's strategic objectives and the types of IT infrastructure services required to support those objectives (Broadbent et al. 1999). TMT with higher absorptive capacity will be better able to align organization-wide IT infrastructure investments with business strategies. This reasoning leads to the following hypothesis:

H5: TMT's absorptive capacity will positively enhance the IT infrastructure sophistication of the organization.

Control Variables

Prior studies suggest that two ancillary factors can influence assimilation: time since adoption (Anderson and Young 1999) and firm size (Davila and Foster 2005, 2007). Two proxies are used for firm size: Number of employees and gross revenue of the firm (Zhu and Kemerer 2002; Subramani 2004; Liang et al. 2007). These factors are modeled as control variables to isolate primary influences on BI assimilation.

III. RESEARCH METHOD

We use a field survey method to test the research hypotheses. The survey was distributed to Australian client organizations of a single international vendor providing BI software. The vendor provides a major BI system used internationally and that is specifically recognized for its packaged MCS capability. The system is configurable for leading ERP packages, enabling the BI system to directly access the underlying enterprise databases. There are over 200 different pre-defined standard reports using over 500 key performance indicators (KPIs) and containing analytics to answer over 2900 business-critical questions. The metrics and reports support MCS capabilities for sales analysis, financial analysis, inventory and procurement analysis, and supply chain analysis, including a broad range of scorecards for assessing performance (Howard 2003). Focusing on a single software vendor controls for variation that may occur from differences in MCS capability across BI software options.

We distributed surveys to 1,873 managers in 612 organizations that the BI vendor provided (subject to a written nondisclosure agreement) as a contact list of clients. Where possible, we selected multiple respondents for each organization from the vendor's contact list to include senior executives, operational managers, and IT users. For a "small organization" that only had a single contact person, the organization was selected if the contact person was a senior executive (e.g., chief executive officer [CEO], chief financial officer [CFO], or chief information officer [CIO]). A multiple-respondent strategy is preferred for the richness of the data, to mitigate bias, and to enhance accuracy (Sethi and King 1994; Huber and Power 1985).

The survey protocol followed the guidance of Dillman (2000). We mailed survey packets including a cover letter, survey, and pre-paid reply envelope to each selected contact. A first reminder was sent by email four weeks later to all the recipients. A second survey packet was sent four weeks after the email reminder to all nonrespondents. A final reminder was emailed two weeks later with a URL link to the web-based version of the survey. Online surveys are used in prior studies for both sole and supplementary survey methods (Dillman 2000). Our tests reveal no differences between paper and online responses.

An average of three respondents in each targeted organization received the survey. We asked respondents from organizations with multiple strategic business units (SBUs) to choose whether to answer the survey on behalf of either a SBU or the whole enterprise. We received a total of 436 responses from 229 organizations including 65 online responses. Due to missing data and selection of "No Basis for Answering" responses on the main study variables, 17 responses are deemed unusable. This resulted in 419 usable responses, for a usable response rate of 22 percent and 36 percent for individual and organization response, respectively.[9] The responses included 135 choosing to identify and respond on behalf of their SBUs. Including responses for SBUs and responses for organizations as a whole, the final sample consists of 347 organizational units.[10]

The average age of respondents is 41.1 years, with 16.5 years of work experience. Respondents are 80 percent male and 20 percent female. Respondents classifying themselves as business executives/managers comprise 54 percent, with 46 percent IT executives/managers and 13 percent holding both business and IT jobs. Over half (54 percent) of respondents report five to eight years of BI systems experience, while 26 percent have over eight years of experience. Large organizations are the predominant respondents, with an average of 663 employees and gross revenue exceeding AUD$2 billion. The sample is also diverse in industry representation, with the most common being manufacturing (65, 18.7 percent) and Retail/Wholesale/Distribution (50, 14.4 percent). Double-digit responses were also received in Banking/Finance/Insurance, Transport/Logistics, Media/Entertainment/Publishing, Healthcare, Telecom, Agricultural/Mining/Construction, and Consulting/Professional Services.

To test the consistency of the responses, we compute the correlation of multiple responses from the same organization on the main constructs (see Armstrong and Sambamurthy [1999] for details). All correlations between two or more respondents from the same organization on the main constructs are significant ($p < 0.01$). The results provide strong evidence of the consistency between responses from a single organization. As a consequence, we use the average scores from multiple respondents as the organizational response. For organizations with a single respondent, the individual response is used to represent the organization.[11]

We perform three tests for common method bias: Harman's one-factor, partialling out a "marker variable," and partialling out a general factor score (Podsakoff and Organ 1986; Podsakoff et al. 2003).[12] The results indicate no significant common method variance that threatens the quality of the data.

---

[9] In addition, 91 respondents reported that they were not the correct informants to answer the survey. Another 70 sent their apology by email and often quoted reasons for not responding to the survey (such as company policy).

[10] Early and late responses were compared in paired samples of 150, 100, 50, and 40 using an ANOVA test for nonresponse bias. The results show no significant differences on any of the study variables, including demographic and control variables. There is no indication of any nonresponse bias.

[11] An ANOVA test shows no significant differences between the individual and averaged responses.

[12] Harman's one-factor test showed neither a single factor emerged from the exploratory factor analysis nor did one general factor account for the majority of the variance in the measurement items used in the model. In partialling out an unrelated "marker variable" as a surrogate for common variance, we used theoretically unrelated variables (respondents' age and years of experience) and examined the structural model both with and without the "marker variables." The results suggest that neither of the two marker variables is statistically significant. In performing the *partialling out a general factor score* test, we added the highest factor from the "unrotated" exploratory factor analysis test to the PLS model as a control variable on the dependent and mediating variables in the model. It is assumed that this factor contains

Operationalization of Constructs

BI Assimilation

BI Assimilation is measured by adapting Armstrong and Sambamurthy's (1999) instrument for IT assimilation. Building on Porter's (1985) value chain framework, Armstrong and Sambamurthy (1999) use 14 items to measure IT assimilation, six that capture IT assimilation in different business activities and eight items that measure IT assimilation in business generic strategies. Armstrong and Sambamurthy (1999) report three dimensions for IT assimilation: logistics activities, marketing activities, and business strategies. A comparison of factors in Armstrong and Sambamurthy's (1999) measure with those identified by the literature as representative of advanced MCS suggest that most attributes involving enterprise data are captured. One item, "managerial processes," is added based on BI literature and feedback in a focus group meeting. This process results in a total of 15 items to measure BI assimilation.

Exploratory factor analysis supports the use of only 14 items. We eliminate one item from the measurement list due to high cross-loadings (more than or closer to the threshold of 0.50).[13] The result from the exploratory factor analysis shows the remaining 14 items load on three factors of BI assimilation. These factors are referred to as (1) *customer relations*, (2) *business operations*, and (3) *marketing and sales*. Each factor combines both generic strategies and business activities related to a specific business function.

Organizational Absorptive Capacity

The measures of absorptive capacity at the TMT and operational levels in this study reflect the two core elements of the absorptive capacity: *prior relevant knowledge* and *intensity of effort* (Cohen and Levinthal 1990; Kim 1998). We use the four modes of knowledge creation suggested by Nonaka (1994): socialization, externalization, combination, and internalization[14] as the basis to measure *intensity of effort* for both operational and TMT levels of the organization. The four knowledge-creation modes represent the organization's effort and ability to create and absorb new information from internal and external sources, convert it into new usable knowledge, and apply it to support strategic planning and execution of business strategies (Nonaka 1994; Davenport and Prusak 1998).

Operational-Level Absorptive Capacity

The measure of *intensity of effort* at the operational level consists of 25 items divided between socialization (seven items), externalization (six items), combination (five items), and internalization (seven items).[15] Confirmatory factor analysis supports the use of only 22 items to measure the

---

the best approximation of the common method variance (Podsakoff and Organ 1986; Podsakoff et al. 2003). The findings show the original results are not affected by the general factor included in the model.

[13] The new item "managerial processes" was removed from the measure of BI assimilation as it cross-loaded on to two factors. One plausible explanation for the cross-loading is that the item refers to broad managerial activities that span the whole value chain. Note also that our three dimensions of BI assimilation do not exactly align with those of Armstrong and Sambamurthy (1999). Our focus is on the whole value chain, while Armstrong and Sambamurthy (1999) tested the assimilation of "IT in general." However, all of their items do load on the overall construct and the variance in sub-categorization is likely due to the diverse nature of the industries and the different type of IT used in our study.

[14] Nonaka (1994) and Nonaka and Takeuchi (1995) suggest that organizational knowledge creation is captured by four modes of knowledge conversion: *socialization* (the process of creating tacit knowledge through shared experience), *externalization* (the process of converting tacit knowledge into explicit knowledge), *combination* (the process of creating explicit knowledge from explicit knowledge), and *internalization* (the process of converting explicit knowledge into tacit knowledge). Each of the four modes creates new knowledge of a specific type, and the dynamic between the four modes enables organizational knowledge creation.

[15] The derived measure of intensity of effort is based on three prior studies (Nonaka et al. 1994; Becerra-Fernandez and Sabherwal 2001; Choi and Lee 2002) that used Nonaka's (1994) four modes of knowledge creation. However, the three studies were conducted in three different countries (Japan, U.S., and South Korea) and the measurement items used in

four modes of knowledge creation. We eliminate three items from the operational-level, knowledge-creation measure because they load below 0.50 on the relevant dimension (Hair et al. 1998). We then use measurement indices produced by Smart PLS to examine the properties of measures for the knowledge-creation modes.[16] All four measures show satisfactory levels of internal consistency and convergent and discriminant validity.[17] We derive four composite variables from average scores corresponding to the four modes to proxy for intensity of effort by averaging the respondents' scores.

The *prior relevant knowledge* component of absorptive capacity captures the shared knowledge between line and IT managers. Nelson and Cooprider (1996) define shared knowledge as the *understanding* and *appreciation* among IT and line managers for the technologies and processes that affect their mutual performance. We use the five items suggested by Nelson and Cooprider (1996) to measure shared knowledge between line and IT managers, which represent two types of measures, multiplicative and general.

(1) For *multiplicative* or interaction measures, respondents are asked to assess separately the role of IS and line managers for two characteristics: Items 1 and 2 capture understanding, while items 3 and 4 capture appreciation. Using the conceptualization of fit interaction (Venkatraman 1989; Nelson and Cooprider 1996), we operationalize the two concepts of understanding and appreciation by multiplying the two relevant items for each of the concepts (i.e., item1 $\times$ item2 and item3 $\times$ item4).

(2) For *general* measures, we ask respondents to rate the overall level of appreciation line managers and IS managers have for each other's accomplishments (item 5).

This procedure results in three items to capture shared knowledge (one general item and two from the outcome of the multiplication). The multiplicative measures provide stronger evidence of the validity of the measurement instrument than would be possible from only one type of measure (e.g., general; Nelson and Cooprider 1996). This is because the distribution of the final measure score depends on the extent to which two indicators agree with each other.

We derive the operational-level, absorptive-capacity score by creating the interaction term between the intensity of effort and shared knowledge variables. We use the two-step score construction procedure suggested by Chin et al. (2003) to estimate the composite variable of the interaction term because the measurement of one of the interaction variables (intensity of effort) is modeled formatively (Chin et al. 2003; Lu and Wang 2008).[18]

---

these three studies are not the same. One concern was that some items used could be driven by country-specific culture that may not be relevant to Australian organizations. To overcome this concern, we synthesized the knowledge-creation modes measures reported in the three papers. To improve measure validity, items in common to at least two of the three instruments were chosen for the initial measurement draft. The initial measurement draft was subjected to a number of pilot-tests including interviews with senior managers, feedback from academics, two focus group meetings, and a small survey. As a result of the pilot testing, seven additional items were added to the initial measurement list.

[16] We use Smart PLS to examine the measurement properties of the two measures (i.e., at both operational and TMT levels) of Nonaka's (1994) knowledge-creation modes using the two-stage method. The eight knowledge-creation modes (4 modes to proxy for TMT's intensity of effort and 4 to proxy for operational managers' intensity of effort) were concurrently examined in one model. The results indicate that each of the measurement modes shows high convergent and discriminant validity.

[17] The results indicate the following for all measures: composite reliability measures are above 0.70, the average variance extracted is above 0.50, items load higher on their associated factors than on any other factors, and the square root of AVE of all constructs is higher than its correlation with any other construct in the model.

[18] As Chin et al. (2003, 11) explain, the formative indicators are not assumed to reflect the same underlying construct (i.e., can be independent of one another and measuring different factors). Therefore, the product indicators (in our model the product of multiplying the four scores for intensity of effort by the three scores for shared knowledge) will not necessarily tap into the same underlying interaction effect. Following the two-step procedure, the first step entails using

TMT's Absorptive Capacity

For the intensity of effort element of absorptive capacity, we develop draft measures for Nonaka's (1994) four knowledge-creation modes using prior literature on TMT (Hambrick 1981, 1995; Hurst et al. 1989; Kakabadse et al. 1995; Kippenberger 1997; Wu et al. 2002). We include 31 items in the initial measurement draft. We discussed the initial measurement items with two senior managers and a knowledge management consultant. We ask the three managers in three separate meetings (1) whether the items included in the list represent the knowledge-creation activities that usually take place at the TMT level, and (2) whether there are any additional knowledge-creation activities relating to the TMT that should be included. This process was followed by several pilot tests including feedback from academics, two focus group meetings, and a small survey. Feedback from participants during the pilot tests was analyzed and incorporated in the measurement draft when appropriate. We include 26 items in the final survey, representing the four modes: socialization (seven items), externalization (six items), combination (six items), and internalization (seven items).

Exploratory factor analysis supports only 19 of the items for measuring knowledge creation at the TMT level. We eliminate seven items because they load below 0.50 on the relevant factor (Hair et al. 1998). All measures of the four knowledge-creation modes show satisfactory levels of internal consistency and convergent and discriminant validity.

We create four composite variables representing the four knowledge-creation modes by averaging the respondents' scores for each mode.[19] The prior relevant knowledge component of absorptive capacity is measured using three items suggested by Armstrong and Sambamurthy (1999) to capture TMT's strategic knowledge of IT. The results reported in Table 1 indicate that the measure of TMT's strategic knowledge has high internal consistency and convergent and discriminant validity. We again follow Chin et al.'s (2003) two-step approach to create the TMT absorptive capacity composite scores using TMT's strategic knowledge and intensity of effort composite variables.

IT Infrastructure Sophistication

We measure IT infrastructure sophistication using a scale adapted from Armstrong and Sambamurthy (1999), refining the scale by drawing on literature from BI, interviews with BI professionals, and meetings with two focus groups. We assess the level of *IT infrastructure sophistication* by asking respondents to indicate the extent to which their organization had *diffused* the ten IT infrastructure components (see Table 1). The final measurement instrument consists of ten key IT infrastructure components that capture two dimensions of enterprise systems related IT infrastructure (i.e., generic and specialized IT infrastructure) and combine to provide the basis for a shared data environment.

---

the formative indicators in conjunction with PLS to create the underlying construct scores for predictor and moderator variables. In step two we use the two composite variables to create a single interaction term representing the absorptive capacity variable.

[19] We conducted sensitivity tests by creating composite variables for each knowledge-creation mode (at both operational and TMT level) using weighted averages whereby each item load is multiplied by the corresponding measure score. We added the outcomes and calculated the average as the composite variables (modes) score. Results remain the same. Measurement properties were discussed in footnote 14.

TABLE 1

Individual Item Loadings, Composite Reliability, Average Variance Extracted (AVE) Statistics

Panel A: Assimilation—Customer Relation: (Composite Reliability = 0.92, AVE = 0.61)

	Loading	Standard Error	t-statistics[a]
ASA1: Customer services (e.g., improving customer satisfaction).	0.77	0.03	29.34
ASA2: Delivery of products/services.	0.77	0.03	26.14
AST3: Enhancing customer relations.	0.75	0.03	22.68
AST5: Enhancing existing products/services.	0.80	0.02	32.31
AST6: Providing value-added goods/services to customers.	0.85	0.02	49.37
AST7: Creating new products/services.	0.73	0.03	26.53

Panel B: Assimilation—Business Operations (Composite Reliability = 0.88, AVE = 0.60)

	Loading	Standard Error	t-statistics[a]
ASA5: Supplier management (e.g., inbound logistics or purchasing).	0.83	0.02	48.18
ASA7: Manufacturing and/or internal operations.	0.74	0.04	21.08
AST1: Being a low-cost producer/provider.	0.69	0.04	17.58
AST2: Creating flexible manufacturing/operations processes.	0.76	0.03	27.40
AST4: Enhancing supplier relations.	0.83	0.02	48.07

Panel C: Assimilation—Marketing and Sales (Composite Reliability = 0.87, AVE = 0.64)

	Loading	Standard Error	t-statistics[a]
ASA3: Marketing (e.g., targeting customers and tailoring offers).	0.84	0.02	35.85
ASA4: Sales (e.g., sales force automation, revenue management).	0.84	0.02	41.81
AST8: Entering new markets.	0.80	0.02	35.73

(continued on next page)

Panel D: Operational Managers' Shared Knowledge: (Composite Reliability = 0.95, AVE = 0.86)

	Loading	Standard Error	t-statistics[a]
Shared1*2: Line managers (Information systems manager) understand the work environment (problems, tasks, roles, etc.) of the information systems managers (line managers).	0.89	0.02	60.36
Shared3*4: Line managers (information systems managers) appreciate the accomplishments of the information systems managers (line managers).	0.95	0.01	162.48
Shared5: Information systems managers and line managers appreciate each others' accomplishments.	0.94	0.01	131.57

Panel E: TMT Knowledge: (Composite Reliability = 0.92, AVE = 0.80)

	Loading	Standard Error	t-statistics[a]
BEK1: The potential and limitations of the organization's…	0.89	0.01	61.84
BEK2: The potential and limitations of emerging information…	0.93	0.01	107.65
BEK3: How competitors are applying information technologies…	0.86	0.02	41.62

Panel F: Generic-IT Infrastructure Sophistication (Composite Reliability = 0.86, AVE = 0.52)

	Loading	Standard Error	t-statistics[a]
ICT1: Mainframe/server systems.	0.62	0.05	11.74
ICT2: Database/ERP systems.	0.56	0.05	11.51
ICT7: Internal computer network.	0.71	0.04	16.08
ICT8: IT architecture and standards.	0.77	0.02	33.83
ICT9: Security and risk management policies.	0.84	0.02	44.20
ICT10: The latest back-end technology.	0.77	0.02	31.18

Panel G: Specialized-IT Infrastructure Sophistication (Composite Reliability = 0.87, AVE = 0.64)

	Loading	Standard Error	t-statistics[a]
ICT3: Data warehouse/data marts.	0.86	0.02	50.74
ICT4: Extract, Transform, Load (ETL) tools.	0.88	0.01	59.02
ICT5: Analytical/reporting tools.	0.77	0.03	23.88
ICT6: IT that enable electronic access to external data.	0.66	0.04	15.50

(continued on next page)

Panel H: TMT-level Knowledge Creation Modes used (proxy for Intensity of Effort at the TMT Level): Formative Construct

	Indicator Weight	VIF	Standard Error	t-statistics
SOC-TMT: Socialization	0.17	1.868	0.033	5.13
EXT-TMT: Externalization	0.26	2.782	0.040	6.44
COM-TMT: Combination	0.28	2.313	0.038	7.38
INT-TMT: Internalization	0.47	2.447	0.041	11.47

Panel I: Operational-Level Knowledge Creation Modes (Proxy for Intensity of Effort at Operational Level): Formative Construct

	Indicator Weight	VIF	Standard Error	t-statistics
SOC-OP: Socialization	0.29	1.979	0.044	6.63
EXT-OP: Externalization	0.19	1.812	0.046	4.16
COM-OP: Combination	0.27	2.086	0.044	6.14
INT-OP: Internalization	0.45	2.482	0.049	9.13

[a] All t-statistics reported in this table are statistically significant at $p < 0.001$.

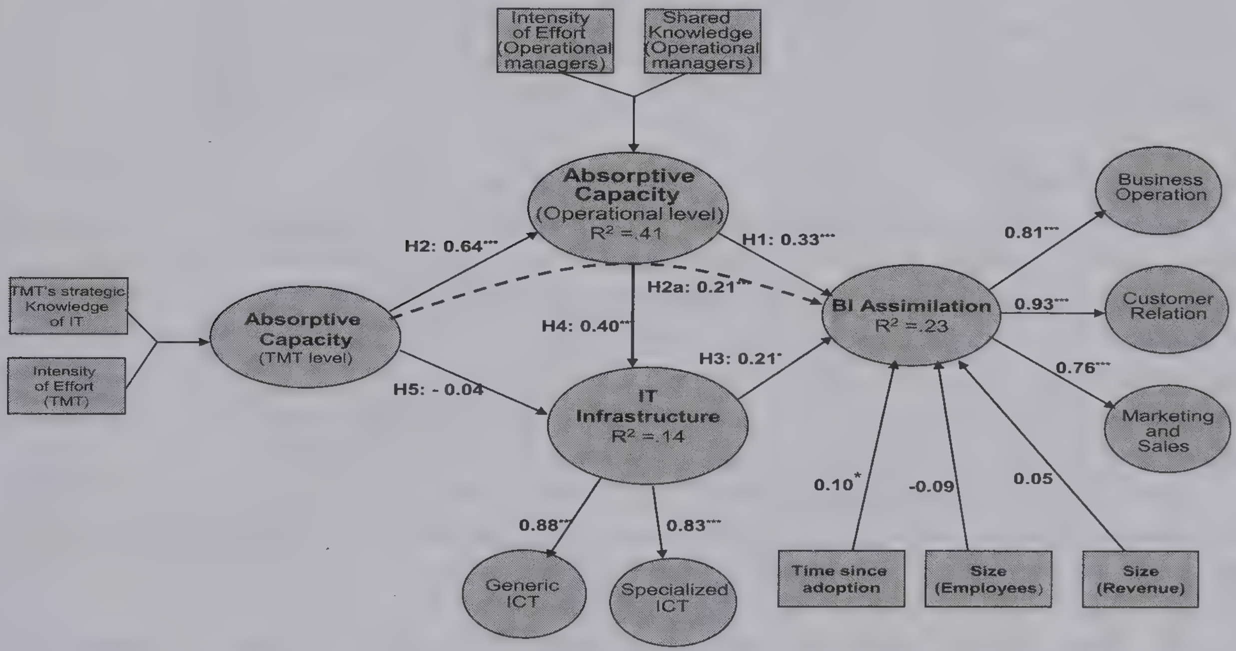
IV. RESULTS

We use partial least squares (PLS), a component-based structural equation modeling technique, to validate our constructs' measures and test the research model and hypotheses.[20] The overall results are shown on the structural model presented in Figure 2.

Measurement Properties

We apply multiple tests suggested by Churchill (1979) and Straub (1989) to assess construct validity and reliability. First, we use either *exploratory* or *confirmatory* factor analyses, depending on the maturity of the measure, to examine the dimensions and loading of measurement items. Second, we simultaneously test the structural model and the measurement model using PLS. The output from PLS in relation to the measurement model verifies the initial results from the exploratory factor analysis tests. Item loadings measure the significance of the item to the factor. For constructs measured with reflective indicators, we drop items with loadings below 0.70 from the construct measure as they indicate that less than 50 percent of the variance of the variable is

FIGURE 2
Structural Model Results for BI Assimilation



*, **, *** Indicate that the coefficient is significant at the $p < .05$, $p < .01$, and $p < .001$ levels, respectively.

[20] We use Smart PLS 2.0. PLS is a component-based structural equation modeling (SEM) technique that simultaneously tests the psychometric properties of the scales used to measure the constructs (i.e., measurement model) and examines the strength of the relations between the constructs (i.e., structural model) (Chin 1998; Hulland 1999). We use PLS instead of covariance-based SEM techniques (such as LISREL) because some of the constructs included in the research model are measured using formative indicators and using covariance-based modeling technique to test the research model can result in an unidentified model (Kline 2006). PLS is also suitable for analyzing complex models with mediating constructs and second-order constructs (Chin and Newsted 1999). A number of recent accounting research studies use PLS for similar reasons (e.g., Hall 2008; Dowling 2009; Chapman and Kihn 2009).

accounted for by the factor (Hair et al. 1998).[21] However, we use a less strict rule of thumb for newly introduced measurement items (Chin 1998; Hulland 1999). To test formative construct validity we assess the indicators' weight rather than the loadings.

Construct Validity

Convergent Validity

Item loading, together with the average variance extracted (AVE), captures the convergent validity of each of the measures for constructs that are modeled reflectively (Van den Bosch 1999). Table 1 shows all reflective measurement items have high and significant loadings, and all the weights of items in formative constructs are statistically significant indicating their significant contribution to the measured construct. The AVE for all constructs exceed 0.50 (ranging between 0.52 and 0.86), supporting the convergent validity of the measurement items (Fornell and Larcker 1981).

Discriminant Validity

Table 2 shows that the values of the square roots of the AVE are all greater than the inter-construct correlations. This indicates that all measures have appropriate discriminant validity. An additional test of discriminant validity assesses each measurement item to ensure that it has a higher loading on its assigned factor than on the other factors (Chin 1998; Gefen et al. 2000) (see Table 3). Each measurement item loads higher on the appropriate construct than on any other construct (Chin 1998; Gefen et al. 2000), providing additional support as to the discriminant validity of the measures.

Hypotheses Testing

We use Smart PLS 2.0 with bootstrapping as a resampling technique (1000 random samples) to estimate the structural model and the significance of the paths. We use path coefficients and the R^2 jointly to evaluate the model (Chin 1998). Table 4 presents the PLS analysis results. As shown in Table 4, of the three control variables included in the research model (revenue and number of employees representing firm size in different ways, and time since adoption), only time since adoption is significantly related to BI assimilation. This indicates that organizations with more BI experience are able to assimilate BI better than those with less BI experience.

Overall, the results suggest the model has good predictability. The coefficients for all paths in the model, except one, are statistically significant at the 0.05 level. The results also indicate that 23 percent of BI assimilation, 41 percent of operational-level absorptive capacity, and 14 percent of IT infrastructure sophistication are explained by the model.

We test hypotheses within the structural equation model shown in Figure 2. Hypotheses that posit direct relations between constructs (H1, H2, H3, H4, and H5) were tested based on the magnitude and significance of path coefficients estimated using Smart PLS 2.0. We tested the hypothesis that posits a mediated (indirect) relation (H2a) by calculating the magnitude and sig-

[21] The measurement items of BI assimilation, shared knowledge at operational level, and TMT's strategic knowledge, knowledge-creation modes (at both operational and TMT levels), and IT infrastructure sophistication are modeled reflectively, while the measurement items of intensity of effort (at both operational and TMT levels) are modeled formatively. The categorization of a construct as a formative or reflective is not always clear-cut and is influenced by the researcher judgment (Chwelos et al. 2001; Dowling 2009). We did additional tests similar to those reported in prior studies (Chwelos et al. 2001; Dowling 2009) whereby we examined different alternative measurement models, modeling all constructs in the model as formative, all the constructs as reflective, and various models with different constructs modeled as formative or reflective. All these tests show results similar to those reported in this paper as no paths gained or lost statistical significance or changed in sign. This suggests the results are not driven by how the constructs are modeled (i.e., formative or reflective).

TABLE 2
Inter-Construct Correlations and Square Root of Average Variance Extracted Statistics[a]
(n = 347)

	Constructs Measured Using Reflective Items[b]							Constructs Measured Using Formative Items and Control Variables Measured with One Item				
	1	2	3	4	5	6	7	8	9	10	11	12
(1) Customer Relation (BI Assimilation)	0.79											
(2) Business Operations (BI Assimilation)	0.60	0.77										
(3) Marketing and Sales (BI Assimilation)	0.64	0.40	0.83									
(4) Generic ICT (IT Infrastructure Sophistication)	0.15	0.20	0.09	0.72								
(5) Specialized ICT (IT Infrastructure Sophistication)	0.42	0.29	0.29	0.47	0.80							
(6) Operational Managers' Shared Knowledge	0.25	0.22	0.30	0.24	0.18	0.93						
(7) TMT Knowledge	0.22	0.18	0.18	0.10	0.09	0.48	0.89					
(8) Intensity of Effort (Operational-Managers level)	0.37	0.41	0.33	0.37	0.33	0.54	0.44	1.00				
(9) Intensity of Effort (TMT level)	0.26	0.25	0.28	0.22	0.23	0.56	0.58	0.67	1.00			
(10) Firm Size (No. of Employees)	−0.08	−0.01	−0.05	0.17	0.17	−0.21	−0.18	−0.10	−0.15	1.00		
(11) Firm Size (Gross Revenue)	−0.03	0.02	0.04	0.12	0.17	−0.20	−0.19	−0.06	−0.07	0.67	1.00	
(12) Time Since Adoption	0.08	0.15	0.17	0.14	0.10	−0.03	0.07	0.15	0.06	0.16	0.22	1.00

[a] Diagonal elements are the square roots of the average variance extracted statistics. Off-diagonal elements are the correlations between the latent variables calculated in PLS.
[b] AVE will only be suitable to use when the construct is measured with reflective indicators (i.e., constructs 1 to 7).

TABLE 3
Items Loading and Cross Loading[a]

	Constructs Measured Using Reflective Items						
Items	Customer Relation	Business Operation	Marketing and Sales	Generic Infra-structure	Specialized Infra-structure	Shared Knowledge (Operational)	TMT-Knowledge
ASA1	0.77	0.45	0.47	0.17	0.30	0.18	0.17
ASA2	0.77	0.56	0.38	0.15	0.33	0.12	0.17
AST3	0.75	0.46	0.53	0.10	0.34	0.19	0.08
AST5	0.80	0.44	0.51	0.10	0.36	0.16	0.15
AST6	0.85	0.48	0.51	0.09	0.40	0.30	0.19
AST7	0.73	0.43	0.58	0.11	0.24	0.22	0.27
ASA5	0.49	0.83	0.30	0.16	0.17	0.12	0.17
ASA7	0.34	0.74	0.29	0.13	0.12	0.15	0.12
AST1	0.38	0.69	0.20	0.24	0.29	0.19	0.11
AST2	0.48	0.76	0.39	0.13	0.28	0.21	0.14
AST4	0.58	0.83	0.36	0.13	0.25	0.18	0.14
ASA3	0.52	0.26	0.84	0.05	0.25	0.18	0.09
ASA4	0.51	0.32	0.84	0.09	0.28	0.26	0.12
AST8	0.55	0.41	0.80	0.09	0.19	0.30	0.22
ICT1	0.05	0.18	0.05	0.62	0.22	0.11	0.03
ICT2	0.08	0.21	0.04	0.56	0.34	0.05	0.06
ICT7	0.06	0.02	0.07	0.71	0.25	0.10	−0.03
ICT8	0.14	0.13	0.09	0.77	0.38	0.22	0.10
ICT9	0.09	0.16	0.04	0.84	0.36	0.22	0.13
ICT10	0.22	0.17	0.10	0.77	0.45	0.27	0.13
ICT3	0.36	0.23	0.22	0.38	0.86	0.14	0.12
ICT4	0.35	0.23	0.20	0.38	0.88	0.16	0.12
ICT5	0.37	0.18	0.32	0.31	0.77	0.15	0.03
ICT6	0.25	0.27	0.20	0.43	0.66	0.11	0.00
Shared1*2	0.21	0.18	0.30	0.25	0.19	0.89	0.46
Shared3*4	0.25	0.22	0.27	0.20	0.15	0.95	0.44
Shared5	0.25	0.21	0.26	0.22	0.16	0.94	0.43
BEK1	0.15	0.13	0.15	0.11	0.04	0.42	0.89
BEK2	0.15	0.16	0.15	0.07	0.10	0.45	0.93
BEK3	0.29	0.19	0.17	0.10	0.09	0.41	0.86
SOCO	0.35	0.39	0.23	0.32	0.27	0.42	0.33
EXTO	0.38	0.41	0.33	0.25	0.32	0.34	0.35
COMO	0.31	0.36	0.28	0.32	0.33	0.41	0.33
INTO	0.23	0.27	0.28	0.32	0.22	0.52	0.42
SOCS	0.24	0.26	0.22	0.21	0.21	0.36	0.35
EXTS	0.20	0.20	0.21	0.24	0.15	0.53	0.48
COMS	0.32	0.27	0.38	0.19	0.29	0.47	0.47
INTS	0.17	0.16	0.17	0.14	0.15	0.50	0.58
FIRMSIZRE	−0.03	0.02	0.04	0.12	0.17	−0.20	−0.19
FIRMSIZEM	−0.08	−0.01	−0.05	0.17	0.17	−0.21	−0.18
YEAR	0.08	0.15	0.17	0.14	0.10	−0.03	0.07
Composite Reliability	0.92	0.88	0.87	0.86	0.87	0.95	0.92
AVE	0.61	0.60	0.64	0.52	0.64	0.86	0.82

(continued on next page)

Items	Constructs Measured Using Formative Items OR with One Item (Control Variables)				
	Intensity of Effort (Operational)	**Intensity of Effort (TMT)**	**Size (Revenue)**	**Size (Employees)**	**Time Since Adoption**
ASA1	0.30	0.17	−0.03	−0.06	0.09
ASA2	0.22	0.16	−0.06	−0.07	0.13
AST3	0.26	0.17	0.06	−0.03	0.05
AST5	0.26	0.20	−0.03	−0.04	0.04
AST6	0.40	0.32	−0.04	−0.12	0.04
AST7	0.35	0.26	−0.04	−0.04	0.04
ASA5	0.34	0.21	−0.02	−0.04	0.14
ASA7	0.30	0.10	−0.05	−0.04	0.14
AST1	0.33	0.24	0.11	0.12	0.12
AST2	0.33	0.24	0.02	−0.02	0.06
AST4	0.36	0.21	0.02	−0.03	0.11
ASA3	0.22	0.18	0.07	0.01	0.14
ASA4	0.26	0.22	0.09	−0.02	0.11
AST8	0.35	0.31	−0.05	−0.11	0.17
ICT1	0.10	0.04	0.08	0.06	0.11
ICT2	0.18	0.08	0.16	0.19	0.14
ICT7	0.21	0.16	0.05	0.12	0.03
ICT8	0.36	0.21	0.15	0.18	0.13
ICT9	0.36	0.26	0.07	0.08	0.13
ICT10	0.31	0.19	0.02	0.09	0.06
ICT3	0.32	0.25	0.18	0.16	0.12
ICT4	0.29	0.22	0.16	0.18	0.10
ICT5	0.29	0.19	0.09	0.08	0.03
ICT6	0.21	0.08	0.12	0.13	0.07
Shared1*2	0.50	0.50	−0.17	−0.20	0.02
Shared3*4	0.48	0.54	−0.20	−0.20	−0.04
Shared5	0.48	0.51	−0.19	−0.19	−0.06
BEK1	0.37	0.51	−0.13	−0.16	0.08
BEK2	0.40	0.50	−0.19	−0.19	0.05
BEK3	0.42	0.51	−0.18	−0.13	0.05
SOCO	0.81	0.52	0.00	−0.05	0.17
EXTO	0.79	0.47	0.00	0.00	0.15
COMO	0.82	0.51	−0.08	−0.07	0.13
INTO	0.84	0.65	−0.09	−0.14	0.08
SOCS	0.52	0.77	0.01	−0.05	0.07
EXTS	0.54	0.88	−0.10	−0.17	0.07
COMS	0.58	0.84	−0.06	−0.09	0.07
INTS	0.57	0.87	−0.06	−0.16	0.02
FIRMSIZRE	−0.05	−0.06	1.00	0.67	0.22
FIRMSIZEM	−0.09	−0.14	0.67	1.00	0.16
YEAR	0.16	0.06	0.22	0.16	1.00

(continued on next page)

Items	Constructs Measured Using Formative Items OR with One Item (Control Variables): Intensity of Effort (Operational)	Intensity of Effort (TMT)	Size (Revenue)	Size (Employees)	Time Since Adoption
Composite Reliability	NA[b]	NA[b]	NA[c]	NA[c]	NA[c]
AVE	NA[b]	NA[b]	NA[c]	NA[c]	NA[c]

[a] Cross-loading is obtained by calculating the correlation of the standardized latent variable scores with the standardized value of the item (Pavlou and Gafen 2005).

[b] Composite reliability and AVE will only be suitable to use when the construct is measured with reflective indicators.

[c] Control variables that are measured with one item.

TABLE 4

Path Coefficients: Test and Control Variables

	Path	Path Coefficient	t-statistic (z-score)
Hypotheses Direct Effect	Operational-level AbsCapacity → Assimilation of BI systems (H1)	0.33	6.37***
	TMT's AbsCapacity → Operational-level AbsCapacity (H2)	0.64	18.47***
	IT Infrastructure Sophistication → Assimilation of BI (H3)	0.21	3.47***
	Operational-level absorptive capacity ... IT Infrastructure Sophistication (H4)	0.40	6.87***
	TMT's AbsCapacity → IT Infrastructure Sophistication (H5)	−0.04	0.53
Hypotheses Indirect Effect	TMT'S AbsCapacity → Operational-level AbsCapacity → BI assimilation (H2a)	0.21[a]	(6.10***)
	TMT'S AbsCapacity → Operational-level AbsCapacity → IT Infrastructure Sophistication (*ad hoc* test for H5)	0.25[a]	(6.43***)
Control Variables	Size (No. of employees) → BI assimilation	−0.09	1.61
	Size (Revenue) → BI assimilation	0.05	0.89
	Time since adoption of BI → BI assimilation	0.10	1.98*

* t-statistic > 1.96 is significant at $p < 0.05$ level.

** t-statistic > 2.58 is significant at $p < 0.01$ level.

*** t-statistic > 3.29 is significant at $p < 0.001$ level.

[a] Standardized path.

nificance of mediated paths (Hoyle and Kenny 1999; Subramani 2004). Additional tests examine whether the indirect path represents full or partial mediation (Baron and Kenny 1986; Mathieson et al. 2001).

Hypothesis 1 predicts that increased levels of operational-level absorptive capacity will enhance the levels of BI assimilation. The results shown in Figure 2 and Table 4 support the hypothesis with a strong and significant relationship (0.33, $p < 0.001$). This result supports the belief that operational and line managers' knowledge and knowledge-creation activities (i.e., absorptive capacity) have a positive influence on assimilation.

Hypothesis 2 predicts that increased levels of TMT absorptive capacity enhance the operational-level absorptive capacity. The results shown in Figure 2 and Table 4 again support the hypothesized relationship with a strong and significant relationship (0.64, $p < 0.001$). This finding

provides support for the integrative effects of cultural controls that primarily emanate from TMT-level management and the subsequent impact on organizational learning and knowledge creation at the operational and line levels.

Hypothesis 2a elaborates on this relationship to examine the connection that occurs between the outcome of cultural controls and the core MCS controls provided by BI systems (e.g., planning and cybernetic controls). Hypothesis 2a examines supporting components of this relationship by looking specifically at the relationship between TMT absorptive capacity and BI assimilation with the theorization that this relationship will occur through operational-level absorptive capacity. Using the Hoyle and Kenny (1999) approach to test the mediated relationship, we first look at the indirect effect which is strong and significant (0.21, $p < 0.001$) (see Table 4). We subsequently test whether the mediation is partial or full (Baron and Kenny 1986). Results, as shown in Table 5, confirm full mediation. This provides evidence of the link between cultural controls and other core MCS controls provided by BI, but also reflects the importance of line and operational managers achieving this interrelationship and successful integration of MCS as a package.

Hypothesis 3 predicts that IT infrastructure sophistication will enhance BI assimilation. The results shown in Figure 2 and Table 4 support the hypothesis (0.21, $p < 0.05$). This supports the belief that underlying enterprise systems-related IT infrastructure plays an important role in BI assimilation.

Hypothesis 4 predicts increased levels of operational-level absorptive capacity will enhance levels of IT infrastructure sophistication. The results in Figure 2 and Table 4 show a strong and significant relationship (0.40, $p < 0.001$) supporting the theorized relationship. This finding highlights the key role of operational and line managers in driving development of appropriate IT infrastructure.

Hypothesis 5 is the only hypothesis not supported. Hypothesis 5 predicts that increased levels of TMT absorptive capacity will enhance levels of IT infrastructure sophistication. This finding is surprising, given that several earlier MCS studies report a strong relationship between TMT involvement and successful implementation of specific MCS components. We conduct *ad hoc* tests to determine if this relationship is offset by operational-level absorptive capacity, which we found to be strongly linked to IT infrastructure sophistication. The most likely scenario from a theoretical perspective would be that TMT absorptive capacity influences IT infrastructure sophistication indirectly through operational level absorptive capacity. Tests for this mediation effect, using the same techniques for testing H2a, are significant, and reflect full mediation (see Tables 4 and 5).

While this latter finding is inconsistent with our *a priori* theoretical model, the finding is not

TABLE 5

Nested Model Comparison

Direct Paths	R^2 in Nested Model (No Direct Path)	R^2 in Models with Direct Path	The Magnitude of the Change in $R^2(f^2)$	Pseudo $F = f^2 \times (n - k - 1)$	Conclusion
TMT's absorptive capacity to BI assimilation: An additional test for (H2a)	0.228	0.231	0.004	1.33	Not significant
TMT's absorptive capacity to IT infrastructure: An additional test for (H5)	0.138	0.139	0.001	0.40	Not significant

necessarily inconsistent with an absorptive-capacity view of assimilation. Prior studies focus on TMT involvement and specific MCS implementation, whereas we focus on absorptive capacity and broad-based MCS implementation. Our findings could be interpreted as indicating the critical importance that operational and line managers play in the integration of technology with business processes and the importance at the user level of recognizing where business value can be derived. This finding should be investigated further in future studies.

V. DISCUSSION AND CONCLUSION

Recent research indicates that MCS are evolving to provide a more strategic view of value-chain activities including product development, sales and marketing, customer relations, and performance management (Davila and Foster 2005, 2007). In just the last few years, we have seen a shift in the fundamental role of the management accountant to consistently include internal analysis and risk management activities (DeLoo et al. 2009). This is reflected in the Institute of Management Accountants' (IMA) December 2008 release of a Statement on Management Accounting that changed the definition of a management accountant for the first time in 30 years (IMA 2008, 1):

> Management accounting is a profession that involves partnering in management decision making, devising planning and performance management systems, and providing expertise in financial reporting and control to assist management in the formulation and implementation of an organization's strategy.

In this study, we examine the assimilation of BI systems that automate MCS as a package and support an ever-increasing number of these contemporary responsibilities for management accountants. We examine BI systems provided by one vendor whose product is considered a leader in providing MCS capability. We focus our study on the cultural controls associated with knowledge creation (i.e., absorptive capacity). The resulting organizational absorptive capacity impacts successful assimilation of BI systems.

Our results provide evidence that increased absorptive capacity among operational-level managers is strongly related to both increased levels of BI assimilation and increased levels of sophistication in the underlying technology infrastructure that enables BI systems. At the same time, operational-level absorptive capacity is strongly tied to TMT's absorptive capacity, providing evidence of the importance of cultural controls related to knowledge management and resource development. These results have implications not only for the role of cultural controls in fostering an effective knowledge culture, but also for the view among many researchers that MCS should be considered as a package—i.e., recognizing that specific MCS components are impacted by the surrounding MCS controls as a whole (Chenhall 2003; Malmi and Brown 2008).

The results for BI assimilation provide some evidence of a potential diffusion effect on the management accounting function as the scope of MCS broadens. Several researchers have observed indications of this diffusion effect due to the adoption of enterprise systems and associated increased capability for MCS (see Granlund and Malmi 2002; Caglio 2003; Scapens and Jazayeri 2003; Quattrone and Hopper 2005). Our results indicate that organizations with greater absorptive capacity assimilate BI across business operations, customer relations, and marketing and sales—a phenomenon that could be perceived as showing widespread change consistent with these observations in earlier studies of specific organizations. Future research should explore this diffusion phenomenon in order to better understand the impact on management accountants' roles in organizations and the MCS function overall.

Our research also contributes to understanding the roles of TMT in MCS innovations through an expanded view of organizational absorptive capacity. In particular, we shift the focus of absorptive capacity to a dynamic perspective of knowledge and incorporate a measurement of intensity of effort that has not been considered in prior research on the link between TMT involve-

ment and MCS innovation success. Focusing on this dynamic view of absorptive capacity, we incorporate the knowledge-creation capabilities that are critical to adapting to new information and new technologies. Incorporating this broader measure of absorptive capacity, we are able to provide evidence that both operational-level and TMT absorptive capacity are likely to influence the assimilation of a broad range of strategic systems related to MCS. At the same time, we also show evidence that the absorptive capacities of the operational-level managers are fundamental to establishing the robust IT infrastructure necessary to support strategic systems that support MCS innovations. Operational-level managers' absorptive capacity is similarly key to organizational assimilation of strategic technologies to gain strategic advantage. Prior research has failed to establish the link between TMT's knowledge and assimilation of strategic systems (e.g., Armstrong and Sambamurthy 1999). Our results confirm the TMT effect as indirect and flowing through the operational level. Future research should further explore this indirect effect in order to better understand the mechanisms through which TMT's knowledge can be leveraged through cultural controls. Further, we have focused on one particular aspect of cultural controls, absorptive capacity. Future research should also consider the broader range of cultural controls and the effect on successful assimilation of MCS.

The findings related to absorptive capacity effects on IT infrastructure are also of interest. There is a contrast between operational-level absorptive capacity, which has a significant influence on IT infrastructure sophistication and TMT's absorptive capacity, which does not have a direct effect. *Post hoc* tests reveal an indirect relationship between TMT's absorptive capacity and IT infrastructure with operational-level managers' absorptive capacity fully mediating the relationship. These findings indicate that sophisticated IT infrastructure is developed through socially complex processes that involve collaborations between TMT, IT specialists, and operational-level managers. The cultural controls put in place by TMT have a substantial effect on these collaborations. Moreover, the results also support the calls made in both the accounting and information systems literature for management accountants, line managers, and IT managers to develop broader overlapping knowledge, interpersonal skills, and deeper understanding of the strategic content of the organization (Todd et al. 1995; Nelson and Cooprider 1996; Caglio 2003; Byrd et al. 2004; Dechow and Mouritsen 2005; Quattrone and Hopper 2005; Arnold 2006; Rikhardsson and Kraemmergaard 2006).

In weighing the results of this study, several inherent limitations should be considered. First, although an attempt was made to solicit multiple respondents from each of the targeted firms, only a single contact was available for many of the firms. However, tests show no differences between mean responses received from single respondents versus those from multiple respondents. Second, using the same informants to answer questions on both operational- and strategic-level constructs may result in a respondent's bias or knowledge bias. However, techniques used in the study help to alleviate some of that concern. These include providing respondents with a "No Basis for Answering" option in the survey, capturing the same data from multiple respondents, testing for consistency between multiple respondents, testing for common method variance, and testing for discriminant validity. All of these tests suggest that such concern does not threaten the validity of the results. Future studies could consider using two separate surveys. One would be answered by TMT members and include questions on the TMT characteristics, and the other by middle-/operational-level mangers and include questions on BI assimilation, IT infrastructure, etc. Third, the study investigates TMT knowledge and argues that it enables the cultural controls that are necessary for building organizational absorptive capacity, appropriate IT resources, and supporting BI assimilation. Future research could examine the direct effects of cultural controls on the cybernetic and planning controls that are most enabled by BI systems.

This research has significant practical implication for management accountants, organizational management, managers using MCS, IT consultants, and technology vendors. The clearest

message is that successful provision of MCS capability is not achieved by acquiring "state-of-the-art" software, but arises through developing appropriate organizational capability for generating and using new knowledge to support MCS. This reasoning also emphasizes that technology innovation is driven from the bottom-up by operational managers, as opposed to the dominant view of top-down, top-management-driven innovation. The top-down view is traditionally perceived as the driving force behind MCS capability. The findings from this study also reinforce organizations' continuous effort in knowledge-creation activities at both TMT and operational levels. Organizations should promote knowledge-creation activities as an ongoing process in preference to purposeful and directed short-term knowledge acquisition such as training programs.

REFERENCES

Anderson, S. W., and S. M. Young. 1999. The impact of contextual and process factors on the evaluation of activity-based costing systems. *Accounting, Organizations and Society* 24: 525–559.

Armstrong, C. P., and V. Sambamurthy. 1999. Information technology assimilation in firms: The influence of senior leadership and IT infrastructures. *Information Systems Research* 10 (4): 304–327.

Arnold, V. 2006. Behavioral research opportunities: Understanding the impact of enterprise systems. *International Journal of Accounting Information Systems* 7 (1): 7–17.

Attewell, P. 1992. Technology diffusion and organizational learning: The case of business computing. *Organization Science* 3 (1): 1–19.

Baron, R. M., and D. A. Kenny. 1986. The moderator-mediator variable distinction in social psychological-research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology* 51 (6): 1173–1182.

Bassellier, G., I. Benbasat, and B. H. Reich. 2003. The influence of business managers' IT competence on championing IT. *Information Systems Research* 14: 317–336.

Becerra-Fernandez, I., and R. Sabherwal. 2001. Organizational knowledge management: A contingency perspective. *Journal of Management Information Systems* 18 (1): 23–55.

Bhimani, A. 2003. A study of the emergence of management accounting system ethos and its influence on perceived system success. *Accounting, Organizations and Society* 28: 523–548.

Boynton, A. C., R. W. Zmud, and G. C. Jacobs. 1994. The influence of IT management practice on IT use in large organizations. *Management Information Systems Quarterly* 18 (3): 299–318.

Brignall, S., and J. Ballantine. 2004. Strategic enterprise management systems: New directions for research. *Management Accounting Research* 15: 225–240.

Broadbent, M., P. Weill, and D. Clair. 1999. The implications of information technology infrastructure for business process redesign. *Management Information Systems Quarterly* 23 (2): 159–182.

Byrd, T. A., B. R. Lewis, and D. E. Turner. 2004. The impact of IT personnel skills on IS infrastructure and competitive IS. *Information Resources Management Journal* 17 (2): 38–62.

Byrne, S., and B. Pierce. 2007. Towards a more comprehensive understanding of the roles of management accountants. *European Accounting Review* 16 (3): 469–498.

Caglio, A. 2003. Enterprise resource planning systems and accountants: Towards hybridization? *European Accounting Review* 12 (1): 123–153.

Carte, T. A., A. B. Schwarzkopf, T. M. Shaft, and R. W. Zmud. 2005. Advanced business intelligence at Cardinal Health. *MIS Quarterly Executive* 4 (4): 413–424.

Chapman, C. S. 2005. Not because they are new: Developing the contribution of enterprise resource planning systems to management control research. *Accounting, Organizations and Society* 30: 685–689.

———, and L.-A. Kihn. 2009. Information systems integration, enabling control and performance. *Accounting, Organizations and Society* 34: 151–169.

Chatterjee, D., R. Grewal, and V. Sambamurthy. 2002. Shaping up for e-commerce: Institutional enablers of the organizational assimilation of Web technologies. *Management Information Systems Quarterly* 26 (2): 65–89.

Chenhall, R. H. 2003. Management control systems design within its organizational context: Findings from contingency-based research and directions for the future. *Accounting, Organizations and Society* 28:

127–168.

——— 2005. Integrative strategic performance measurement systems, strategic alignment of manufacturing, learning and strategic outcomes: An exploratory study. *Accounting, Organizations and Society* 30: 395–422.

———, and K. J. Euske. 2007. The role of management control systems in planned organizational change: An analysis of two organizations. *Accounting, Organizations and Society* 32: 601–637.

Chin, W. W. 1998. The partial least squares approach to structural equation modeling. In *Research Methods for Business Research*, 22 (1): edited by G. A. Marcoulides. Mahwah, NJ: Lawrence Erlbaum Associates.

———, and P. Newsted. 1999. Structural equation modeling analysis with small samples using partial least squares. In *Statistical Strategies for Small Sample Research*, 307–341, edited by R. Hoyle. Thousand Oaks, CA: Sage.

———, B. L. Marcolin, and P. R. Newsted. 2003. A partial least squares latent variable modeling approach for measuring interaction effects: Results from a Monte Carlo simulation study and an electronic-mail emotion/adoption study. *Management Information Systems Quarterly* 14 (2): 189–217.

Choi, B., and H. Lee. 2002. Knowledge management strategy and its link to knowledge creation process. *Expert Systems with Applications* 23 (3): 173–187.

Churchill, G. A. 1979. A paradigm for developing better measures of marketing constructs. *JMR, Journal of Marketing Research* 16 (1): 64–73.

Chwelos, P., I. Benbasat, and A. Dexter. 2001. Research report: Empirical test of an EDI adoption model. *Information Systems Research* 12 (3): 304–321.

Cohen, W. M., and D. A. Levinthal. 1990. Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly* 35 (1): 128–152.

Cook, S. D. N., and J. S. Brown. 1999. Bridging epistemologies: The generative dance between organizational knowledge and organizational knowing. *Organization Science* 10 (4): 381–400.

Daft, R. L., and K. E. Weick. 1984. Toward a model of organizations as interpretive systems. *Academy of Management Review* 9: 284–295.

Daghfous, A. 2004. Absorptive capacity and the implementation of knowledge-intensive best practices. *SAM Advanced Management Journal* 69 (2): 21–27.

Davenport, T., and L. Prusak. 1998. *Working Knowledge: How Organizations Manage What They Know.* Boston, MA: Harvard Business School Press.

Davila, A., and G. Foster. 2005. Management accounting systems adoption decisions: Evidence and performance implications from early-stage/startup companies. *The Accounting Review* 80 (4): 1039–1068.

———, and ———. 2007. Management control systems in early-stage startup companies. *The Accounting Review* 82 (4): 907–937.

Dechow, N., and J. Mouritsen. 2005. Enterprise resource planning systems, management control and the quest for integration. *Accounting, Organizations and Society* 30: 691–733.

DeLoo, I. , B. Verstegen, and D. Swagerman. 2009. Towards a better understanding of management accountants' roles in the Netherlands: The management accountants survey 2007. Working paper, Open University of the Netherlands. Available at SSRN: http://ssrn.com/abstract=1368119.

Dillman, D. A. 2000. *Mail and Internet Surveys: The Tailored Design Method.* 2nd edition. New York, NY: John Wiley and Sons, Inc.

Ditillo, A. 2004. Dealing with uncertainty in knowledge-intensive firms: The role of management control systems as knowledge integration mechanisms. *Accounting, Organizations and Society* 29: 401–421.

Dowling, C. 2009. Appropriate audit support system use: The influence of auditor, audit team, and firm factors. *The Accounting Review* 84 (3): 771–810.

Elbashir, M., and S. Williams. 2007. BI impact: The assimilation of business intelligence into core business processes. *Business Intelligence Journal* 12: 45–54.

Fichman, R. G., and C. F. Kemerer. 1997. The assimilation of software process innovations: An organizational learning perspective. *Management Science* 43 (10): 1345–1363.

Fornell, C. D., and F. Larcker. 1981. Evaluating structural equation models with unobservable variables and measurement errors. *JMR, Journal of Marketing Research* 18: 39–50.

Gefen, D., D. W. Straub, and M. Boudreau. 2000. Structural equation modeling and regression: Guidelines

for research practice. *Communications of The Association for Information Systems* 4 (Article 7): 1–77.

Gnatovich, R. 2007. Making a case for business analytics. *Strategic Finance* (February): 47–51.

Granlund, M., and T. Malmi. 2002. Moderate impact of ERPs on management accounting: A lag or permanent outcome? *Management Accounting Research* 13: 299–321.

Grant, R. M. 1996. Toward a knowledge-based theory of the firm. *Strategic Management Journal* 17: 109–122.

Hair, J. F., R. E. Anderson, R. L. Tatham, and W. C. Black. 1998. *Multivariate Data Analysis*. Upper Saddle River, NJ: Prentice Hall.

Hall, M. 2008. The effect of comprehensive performance measurement systems on role clarity, psychological empowerment and managerial performance. *Accounting, Organizations and Society* 33: 141–163.

Hambrick, D. C. 1981. Strategic awareness within top management teams. *Strategic Management Journal* 2 (3): 263–279.

——— 1995. Fragmentation and the other problems CEOs have with their top management teams. *California Management Review* 37 (3): 110–127.

Hambrick, D. C., and P. A. Mason. 1984. Upper echelons: The organization as a reflection of its top managers. *Academy of Management Review* 9 (2): 193–206.

Howard, P. 2003. *Analytics Volume 1: An Evaluation and Comparison*. Edited by B. Smithson. Milton Keynes, U.K.: Bloor Research.

Hoyle, R. H., and D. A. Kenny. 1999. Sample size, reliability, and tests of statistical mediation. In *Statistical Strategies for Small Sample Research*, edited by R. H. Hoyle. Thousand Oaks, CA: Sage Publications.

Huber, G. P., and D. J. Power. 1985. Retrospective reports of strategic-level managers: Guidelines for increasing their accuracy. *Strategic Management Journal* 6 (2): 171–180.

Hulland, J. 1999. Use of partial least squares (PLS) in strategic management research: A review of four recent studies. *Strategic Management Journal* 20 (2): 195–204.

Hurst, D. K., J. C. Rush, and R. E. White. 1989. Top management teams and organizational renewal. *Strategic Management Journal* 10: 87–105.

Institute of Management Accountants. 2008. *Statement on Management Accounting: Definition of Management Accounting*. Montvale, NJ: IMA.

Jansen, J. J. P., F. A. Van den Bosch, and H. W. Volberda. 2005. Managing potential and realized absorptive capacity: How do organizational antecedent's matter? *Academy of Management Journal* 48 (6): 999–1015.

Kakabadse, A., J. McMaho, and A. Myers. 1995. Correlates of internal and external leadership of top management teams: An international comparative study. *Leadership and Organization Development Journal* 16 (7): 10–17.

Kay, M. 2006. The BPM convergence. *Strategic Finance* (September): 51–55.

Keen, P. G. W. 1991. *Shaping the Future: Business Design Through Information Technology*. Boston, MA: Harvard Business School Press.

Kim, L. 1998. Crisis construction and organizational learning: Capability building in catching-up at Hyundai Motor. *Organization Science* 9 (4): 506–521.

Kippenberger, T. 1997. Organizational learning starts at the top. *The Antidote* 2 (7): 19–22.

Kline, R. B. 2006. Reverse arrow dynamics. Formative measurement and feedback loops. In *Structural Equation Modeling: A Second Course*, edited by G. R. Hancock, and R. O. Mueller, 43–68. Charlotte, NC: Information Age Publishing.

Liang, H., N. Saraf, Q. Hu, and Y. Xue. 2007. Assimilation of enterprise systems: The effect of institutional pressures and the mediating role of top management. *Management Information Systems Quarterly* 31 (4): 59–88.

Lu, H.-P., and S. Wang. 2008. The role of internet addiction in online game loyalty: An exploratory study. *Internet Research* 18 (5): 499–519.

Malmi, T., and D. A. Brown. 2008. Management control systems as a package—Opportunities, challenges and research directions. *Management Accounting Research* 19: 287–300.

Mathieson, K., E. Peacock, and W. Chin. 2001. Extending the technology acceptance model: The influence of perceived user resources. *The Data Base for Advances in Information Systems* 32 (3): 86–112.

McCarthy, W. E. 1982. The REA accounting model: A generalized framework for accounting systems in a

shared data environment. *The Accounting Review* 57 (3): 554–578.

Nambisan, S., R. Agarwal, and M. Tanniru. 1999. Organizational mechanisms for enhancing user innovation in information technology. *Management Information Systems Quarterly* 23 (3): 365–395.

Naranjo-Gil, D., and F. Hartmann. 2006. How top management teams use management accounting systems to implement strategy. *Journal of Management Accounting Research* 18: 21–53.

———, and ———. 2007. Management accounting systems, top management team heterogeneity and strategic change. *Accounting, Organizations and Society* 32: 735–756.

Nelson, K. M., and J. G. Cooprider. 1996. The contribution of shared knowledge to IS group performance. *Management Information Systems Quarterly* 20 (4): 409–432.

Nonaka, I. 1994. A dynamic theory of organizational knowledge creation. *Organization Science* 5 (1): 14–37.

———, P. Byosiere, C. C. Borucki, and N. Konno. 1994. Organizational knowledge creation theory: A first comprehensive test. *International Business Review* 3 (4): 337–351.

———, and H. Takeuchi. 1995. *The Knowledge-Creating Company*. New York, NY: Oxford University Press.

———, and N. Konno. 1998. The concept of Ba: Building a foundation for knowledge creation. *California Management Review* 40 (3): 1–15.

———, and R. Toyama. 2003. The knowledge-creating theory revisited: Knowledge creation as a synthesizing process. *Knowledge Management Research and Practice* 1: 2–10.

Pavlou, P. A., and D. Gafen. 2005. Psychological contract violation in online marketplaces: Antecedents, consequences, and moderating roles. *Information Systems Research* 16 (4): 372–399.

Podsakoff, P. M., and D. W. Organ. 1986. Self-reports in organizational research: Problems and prospects. *Journal of Management* 12 (4): 531–544.

Podsakoff, P. M., S. B. MacKenzie, and J.-Y. Lee. 2003. Common method bias in behavioral research: A critical review of the literature and recommended remedies. *The Journal of Applied Psychology* 88 (5): 879–903.

Porter, M. E. 1985. *Competetive Advantage*. New York: Free Press

Quattrone, P., and T. Hopper. 2005. A "time-space odyssey:" Management control systems in two multinational organizations. *Accounting, Organizations and Society* 30: 735–764.

Rikhardsson, P., and P. Kraemmergaard. 2006. Identifying the impacts of enterprise system implementation and use: Examples from Denmark. *International Journal of Accounting Information Systems* 7 (1): 36–49.

Ravichandran, T. 2005. Organizational assimilation of complex technologies: An empirical study of component-based software development. *IEEE Transactions on Engineering Management* 52 (2): 249–268.

Robertson, B., A. Boehler, and J. Hansel. 2007. Sustainable performance improvement through predictive technologies. *Strategic Finance* (June): 57–64.

Rom, A., and C. Rohde. 2007. Management accounting and integrated information systems: A literature review. *International Journal of Accounting Information Systems* 8 (1): 40–68.

Scapens, R. W., and M. Jazayeri. 2003. ERP systems and management accounting change: Opportunities or impacts? A research note. *European Accounting Review* 12 (1): 201–233.

Sethi, V., and W. R. King. 1994. Development of measures to assess the extent to which an information technology application provides competitive advantage. *Management Science* 40 (12): 1601–1627.

Simons, R. 1995. *Levers of Control*. Boston, MA: Harvard University Press.

Straub, D. W. 1989. Validating instruments in MIS research. *Management Information Systems Quarterly* 13 (2): 147–169.

Subramani, M. 2004. How do suppliers benefit from information technology use in supply chain relationships? *Management Information Systems Quarterly* 28 (1): 45–73.

Sutton, S. G. 2000. The changing face of accounting in an IT-driven environment. *International Journal of Accounting Information Systems* 1 (1): 1–8.

Szulanski, G. 1996. Exploring internal stickiness: Impediments to the transfer of best practice within the firm. *Strategic Management Journal* 17: 27–43.

Todd, P. A., J. D. McKeen, and R. B. Gallupe. 1995. The evolution of IS job skills: A content analysis of IS job advertisements from 1970 to 1990. *Management Information Systems Quarterly* 19 (1): 1–27.

Ulrich, D., and M. F. Wiersema. 1989. Gaining strategic and organizational capability in a turbulent business environment. *The Academy of Management Executive* 3 (2): 115–122.

Van den Bosch, F. A. J., H. W. Volberda, and M. de Boer. 1999. Coevolution of firm absorptive capacity and knowledge environment: Organizational forms and combinative capabilities. *Organization Science* 10 (5): 551–568.

Vandenbosch, B. 1999. An empirical analysis of the association between the use of executive support systems and perceived organizational competitiveness. *Accounting, Organizations and Society* 24 (1): 77–92.

Venkatraman, N. 1989. The concept of fit in strategy research: Toward verbal and statistical correspondence. *Academy of Management Review* 14 (3): 423–444.

Vera-Munoz, S. C., J. L. Ho, and C. W. Chow. 2006. Enhancing knowledge sharing in public accounting firms. *Accounting Horizons* 20 (2): 133–156.

von Hippel, E. 1994. Sticky information and the locus of problem solving: Implications for innovation. *Management Science* 40 (4): 429–439.

Williams, S. 2008. Power combination: Business intelligence and the balanced scorecard. *Strategic Finance* 89 (May): 1–8.

——— 2004. Delivering strategic business value: Business intelligence can help management accounting reclaim its relevance and rightful role. *Strategic Finance* 86 (2): 41–48.

———, and N. Williams. 2007. *The Profit Impact of Business Intelligence*. London, U.K.: Elsevier Inc.

Wu, W. Y., C. Y. Chiang, and J. S. Jiang. 2002. Interrelationships between TMT management styles and organizational innovation. *Industrial Management & Data Systems* 102 (3/4): 10–17.

Zahra, S. A., and G. George. 2002. Absorptive capacity: A review, reconceptualization, and extension. *Academy of Management Review* 27 (2): 185–203.

Zhu, K., and K. L. Kraemer. 2002. E-commerce metrics for net-enhanced organizations: Assessing the value of e-commerce to firm performance in the manufacturing Sector. *Information Systems Research* 13 (3): 275–295.

THE ACCOUNTING REVIEW
Vol. 86, No. 1
2011
pp. 185–208

American Accounting Association
DOI: 10.2308/accr.00000013

Disaggregating Management Forecasts to Reduce Investors' Susceptibility to Earnings Fixation

W. Brooke Elliott
Jessen L. Hobson
Kevin E. Jackson
University of Illinois at Urbana–Champaign

ABSTRACT: This study examines disaggregated management forecasts as a mechanism to reduce investors' fixation on announced earnings. Our experimental results suggest that investors' earnings fixation is reduced when they initially observe a disaggregated management forecast (earnings and its components) versus when they observe an aggregated forecast (earnings only). We also provide theory-consistent evidence that this reduction in earnings fixation is associated with investors interpreting the summary net income figure as one of several similarly important evaluation inputs rather than a substantially more important input (relative to its components). Finally, we provide evidence that suggests our results are not bounded by the level of emphasis on net income in the subsequent earnings announcement, and not fully explained by three plausible alternative explanations. Our study extends the voluntary disclosure literature by providing evidence that the form of management disclosures can influence investors' interpretation of subsequently announced information, and contributes to practice by providing a potential alternative to stopping earnings guidance.

Keywords: *earnings fixation; forecast disaggregation; stopping guidance.*

Data Availability: *Data available for reproduction purposes.*

I. INTRODUCTION

Market observers (McCafferty 2007), policy makers (CFA Centre for Financial Market Integrity 2006; Herz 2006), investor advocate groups (CFA Centre for Financial Market Integrity 2006, 2007, 2008) and accounting researchers (e.g., Graham et al. 2005; Chen et al. forthcoming; Cheng et al. 2007; Houston et al. 2010) have each expressed concerns about the

We thank Wendy Bailey, Rob Bloomfield, Annie Farrell, Susan Krische, Laura Li, Bob Libby, Mark Nelson, Mark Peecher, Jane Thayer, Brad Tuttle, Martin Wu, two anonymous reviewers from the American Accounting Association's 2008 ABO Research Conference, and workshop participants at Cornell University, University of Illinois at Urbana–Champaign, University of South Carolina, two anonymous reviewers, and Jim Hunton (associate editor) for helpful comments. We also thank Eric Gooden, Cathy Gray, Ling Harris, Walied Keshk, and Ryan Tate for valuable research assistance, and the financial assistance of the Department of Accounting at Florida State University and the University of Illinois Campus Research Board.

Editor's note: Accepted by James Hunton.

Submitted: May 2009
Accepted: June 2010
Published Online: January 2011

tendency of investors to focus on companies' short-term results in lieu of long-term potential. Often this short-termism is attributed to investors' excessive reliance on accounting earnings as the primary input to investment judgments, without fully considering other information that may be relevant to evaluating the company's investment prospects (Graham et al. 2005; CFA Centre for Financial Market Integrity 2006). Further, this excessive reliance on accounting earnings (i.e., earnings fixation) is often cited as the key source of managers' myopic behavior and associated reduction in long-term value delivered to shareholders (CFA Centre for Financial Market Integrity 2006, 2007).[1]

Although earnings fixation represents an institutional phenomenon that can be undesirable to managers and investors, prior research has primarily focused on documenting the existence and extent of earnings fixation rather than on mechanisms that reduce this fixation. We seek to fill this gap by examining a mechanism to reduce earnings fixation. Specifically, we examine whether and how the level of aggregation of a management forecast can reduce the extent to which investors' investment judgments reflect earnings fixation.

We focus on management forecasts because (1) they are often investors' first exposure to a firm's expectation of its future bottom-line summary earnings, and (2) management has full discretion over the format of the forecast—specifically, the items for which a forecast is provided. Furthermore, several stakeholder groups have advocated the potentially costly approach of eliminating management forecasts to reduce fixation (Chen et al. forthcoming; Houston et al. 2010). We examine a potential alternative to eliminating guidance as a mechanism to reduce earnings fixation: simply altering the level of aggregation in the guidance provided.

We predict that investors' susceptibility to earnings fixation will decrease (increase) when investors initially observe a management forecast that disaggregates (aggregates) the components of forecasted earnings. Our prediction is guided by research in psychology and marketing that suggests that because information can often be perceived in several different ways, the interpretation of that information depends on the particular concept or knowledge structure that is currently active (Higgins and King 1981; Wyer and Srull 1981; Yi 1990a, 1990b).

More specifically, we contend that investors possess at least two general concepts related to net income—one conceptualization of net income as one of several important inputs to evaluating a company's prospects, and another as a sufficient and convenient summary measure, and thus a substantially more important measure, of a company's prospects (relative to its components). We predict that when investors observe a disaggregated earnings forecast, which places net income among a list of other potentially relevant inputs, the former conceptualization will likely be primed. In contrast, when investors observe an aggregated earnings forecast, where net income is isolated, the latter conceptualization of net income will likely be primed. Once the investor observes net income in the subsequent earnings announcement, the primed knowledge structure will be activated and used to evaluate net income. Thus, the investment judgments of investors who initially observe a *disaggregated* (an *aggregated*) earnings forecast are expected to display less (more) earnings fixation because they evaluate net income as one of several similarly important inputs (as a substantially more important input) in assessing the company's prospects.

We use an experiment to test these predictions. Participant-investors start by observing a management earnings forecast and making initial investment judgments. We manipulate the forecast such that it is *disaggregated* or *aggregated*. Consistent with Hirst et al. (2007), disaggregated earnings forecasts are defined as forecasts of the bottom-line net income number as well as

---

[1] The CFA Centre for Financial Market Integrity's (2006) report titled "Breaking the Short-Term Cycle" defines short-termism as the "excessive focus of some corporate leaders, investors, and analysts on short-term, quarterly earnings and a lack of attention to the strategy, fundamentals, and conventional approaches to long-term value creation."

forecasts of each line item on the income statement (e.g., revenues, research and development expenses, etc.). Aggregated earnings forecasts are defined as forecasts of the bottom-line net income number alone. Participants then observe an earnings announcement and make a second set of investment judgments. Although the earnings announcement always reports a net income figure that matches forecasted net income, we manipulate whether the underlying components of net income indicate *more favorable* or *less favorable* performance. This research design allows us to infer earnings fixation when participants' investment judgments are relatively the same regardless of the favorability indicated by the net income components, as such, judgments merely reflect the fact that the announced net income figure is identical across favorability conditions. In contrast, fixation is taken to be reduced when participants' judgments are more (less) favorable when the reported net income components indicate more (less) favorable performance.

As predicted, participants' investment judgments reflect less earnings fixation when they initially observe a disaggregated management earnings forecast. Specifically, participants who initially observe a disaggregated forecast provide investment judgments that are more (less) favorable when the net income components are more (less) favorable. In contrast, participants who initially observe an aggregated forecast provide investment judgments that are insensitive to the reported levels of the net income components. This evidence suggests that the level of aggregation in the initial management earnings forecast influences the extent to which participants' investment judgments reflect line items other than bottom-line net income (i.e., the extent of earnings fixation).

We also find evidence consistent with our theory that this reduction in earnings fixation is associated with investors' interpreting the summary net income figure as one of several similarly relevant evaluation inputs rather than a substantially more important input (relative to its components). We ask participants to allocate 100 points among the bottom-line net income number and its components to indicate each item's importance in evaluating the company as a potential investment. Participants who observe a disaggregated forecast assign less (more) importance to the bottom-line net income number (net income components) than those who observe an aggregated forecast. This result supports our contention that the disaggregated forecast leads investors to consider net income as one of several similarly important inputs in evaluating the company as a potential investment.

In additional analyses, we find evidence that that our results cannot be fully explained by (1) differences in available information across forecast aggregation levels, (2) management signaling precisely which financial statement items are relevant by disaggregating the forecast, or (3) forecast aggregation levels affecting investors' perceived credibility of the forecast. We also find that the effect of a disaggregated forecast is nearly identical regardless of whether the subsequent earnings announcement emphasizes or deemphasizes the bottom-line net income number. Thus, the effect of disaggregating the earnings forecast does not appear to be bounded by the level of emphasis of the bottom-line net income number in the subsequent earnings announcement.

This study makes several contributions to the extant literature. First, we extend accounting research that documents the existence of investors' fixation on earnings (e.g., Bushee 1998) by proposing a potentially attractive approach to reduce fixation in investors' judgments. Second, we extend the disclosure literature in accounting by providing evidence that benchmarks, or forecasts, cannot only be used to evaluate current and prior period information (Schrand and Walther 2000; Krische 2005), but can also affect how investors interpret subsequently reported accounting information. Specifically, we show that a disaggregated earnings forecast can lead investors to interpret subsequently announced net income to be one of several similarly important inputs in evaluating a company's future prospects. Third, we extend the disclosure literature on the effects of forecast characteristics on investors' responses to subsequently announced earnings (Libby and Tan 1999;

Tan et al. 2002; Libby et al. 2006; Han and Tan 2007) by examining the influence of a different characteristic—the level of forecast aggregation—on investors' response to subsequently announced earnings.

This study also extends recent research on the influence of the level of forecast aggregation on investor judgments. Hirst et al. (2007), for example, report results suggesting that investors judge disaggregated forecasts as more credible than aggregated forecasts and deem subsequently reported net income to be of higher quality. One implication is that investors should rely on subsequently reported earnings more when they initially observe a disaggregated forecast. By showing that the level of aggregation influences not only investors' judgments of forecast credibility, but also the extent to which investors' judgments reflect earnings fixation, we refine our understanding of the relationship between forecast aggregation level and investors' reactions to subsequently reported earnings.

Finally, this study contributes to practice. In particular, we provide evidence that modifying the format of management earnings guidance may reduce investors' susceptibility to earnings fixation, while avoiding the potentially costly effects associated with eliminating guidance. This evidence should be important to both managers, who have incentives to reduce pressures for myopic behavior, and regulators, who have expressed concerns about the effects of investors' earnings fixation.

Section II discusses relevant background information and develops our hypothesis. Section III describes our experiment. Section IV presents our main results along with evidence of our theory and describes an alternative theoretical explanation. Section V presents supplemental analyses. Section VI concludes.

II. BACKGROUND AND HYPOTHESIS DEVELOPMENT

Company managers, financial market participants, regulators, and other stakeholders have each raised concerns about the tendency of investors to fixate on earnings (see Bushee 2001; Graham et al. 2005; CFA Centre for Financial Market Integrity 2007). Earnings fixation refers to investors' excessive reliance on a company's accounting earnings in evaluating the firm, without fully considering other information that is relevant to evaluating the company's investment prospects (see Graham et al. 2005).[2] Although its effect on the U.S. economy is not easily quantified, investors' earnings fixation and related short-termism is commonly depicted as detrimental to the judgments and decisions of both managers and investors (Bushee 1998, 2001; Graham et al. 2005).[3]

We argue that earnings fixation occurs when investors conceptualize net income as a substantially more important input (relative to its components) in evaluating a company's prospects, rather than as one of several similarly important inputs. Research in psychology suggests that investors' initial conceptualization of a net income figure is likely to persist when subsequently exposed to the net income figure (Fischhoff 1977). Although announced earnings are undoubtedly important to investors, earnings forecasts often represent investors' initial exposure to a net income amount (*The Economist* 2006; Houston et al. 2010). Thus, our study focuses on earnings forecasts as an opportunity to affect investors' initial conceptualization of net income.

Research in accounting identifies several reasons for companies to issue forecasts. For example, management forecasts help to calibrate analysts' forecasts to minimize analyst forecast errors and forecast dispersion (e.g., Ajinkya and Gift 1984; Houston et al. 2010) and guide

---

2 Our definition of earnings fixation is consistent with the idea of fixating on earnings without regard to how bottom-line net income is derived (Murray 1991).

3 See Graham et al. (2005) for a discussion of managers' myopic decision-making in response to a perceived investor fixation on earnings.

analysts' forecasts toward beatable earnings targets (e.g., Cotter et al. 2006). Management guidance can also reduce litigation risk associated with not warning investors of impending bad news (e.g., Skinner 1994, 1997; Johnson et al. 2001). In addition, issuing forecasts can aid in building a reputation for transparent reporting (e.g., Graham et al. 2005; Hutton and Stocken 2009).

Recently, several stakeholders have advocated the elimination of earnings forecasts, including the CFA Institute and the Business Roundtable, the National Investor Relations Institute, the Aspen Institute's Corporate Values Strategy Group, the Committee for Economic Development, the U.S. Chamber of Commerce, members of Congress, and members of the Securities and Exchange Commission (Janjigian and Ozanian 2006; Houston et al. 2010). Also, several companies have opted to eliminate earnings guidance (e.g., McDonald's Corp. and Merrill Lynch; see also Chen et al. forthcoming; Cheng et al. 2007; Houston et al. 2010). These parties argue that earnings forecasts lead investors to fixate on earnings that, in turn, direct managers to engage in myopic behavior that often runs counter to maximizing the company's long-term growth and shareholder value (Graham et al. 2005; Houston et al. 2010).

Eliminating earnings forecasts could be costly to the company, however. For example, eliminating forecasts may not be a realistic option for smaller companies that are less closely followed by investors (McCafferty 2007). In addition, research in this area finds a negative market reaction and a deterioration of the information environment for firms that stop providing quarterly guidance (Chen et al. forthcoming; Houston et al. 2010). Specifically, Houston et al. (2010) provide evidence of a decrease in analyst coverage of "guidance stoppers," and Bowen et al. (2008) suggest that a decrease in analyst coverage leads to an increase in the cost of raising equity capital.

We propose that, rather than eliminate forecasts, managers might consider changing the type of forecasts they issue.[4] Managers provide forecasts with different levels of aggregation. For example, a recent examination of S&P 500 companies indicates that of the 63 percent of firms that give earnings guidance, 23 percent provide highly aggregated forecasts that forecast earnings alone, while 22 percent provide highly disaggregated forecasts that forecast several components of earnings (see Lansford et al. [2010] and forecast examples at Exhibit 1). This research suggests that managers' choice of whether to issue aggregated or disaggregated forecasts is a function of their desire to signal their ability to forecast, enhance the credibility of earnings forecasts, and respond to external demand for additional information. We argue, however, that the level of managements' forecast aggregation can also affect the extent of investors' earnings fixation.

We predict that a management forecast that is more disaggregated will reduce earnings fixation. Prior research in psychology and marketing suggests that because information can often be perceived in several different ways, the interpretation of that information depends on the particular concept or knowledge structure that is currently active (Higgins and King 1981; Wyer and Srull 1981; Yi 1990a, 1990b). More specifically, because individuals do not undertake an exhaustive memory search for all potential useful concepts, the active or accessible concepts are more likely to be used to process subsequent information (Wyer and Srull 1981; see also Tversky and Kahneman 1973, 1974; Taylor and Fiske 1979).

More closely related to our study, Yi (1990a, 1990b) examined the way an advertising context

[4] We acknowledge that there can also be costs to issuing a disaggregated forecast. For example, providing detailed guidance of a firm's operations may assist competitors. In addition, for firms with complex operations, it may be more difficult for its managers to predict the firm's earnings components (see Lansford et al. 2010 for a discussion). Thus, any given manager must weigh the costs associated with stopping guidance against the costs (benefits) of issuing disaggregated guidance.

EXHIBIT 1

Examples of Management Forecasts

Panel A: Aggregated Management Forecast[a]

XEROX

- Xerox expects first-quarter 2006 earnings in the range of 20-23 cents per share. The company also reiterated its full-year 2006 guidance of $1.00-$1.07 per share. Mulcahy indicated that she now expects the company will deliver full-year earnings in the high end of this range.

Panel B: Disaggregated Management Forecast[b]

intel Leap ahead™

2006 Outlook

- Revenue: Expected to be between $9.1 billion and $9.7 billion.
- Gross margin: 59 percent, plus or minus a couple of points (excluding share–based compensation effects of approximately 1 percent).
- Expenses (R&D plus MG&A): Approximately $3.3 billion (approximately $3 billion excluding share–based compensation effects of approximately $300 million).
- Gains from equity investments and interest and other: Approximately $140 million.
- Tax rate: Approximately 32 percent.
- Depreciation: $1.1 billion, plus or minus $100 million.
- Amortization of acquisition–related intangibles and costs: Approximately $20 million.

[a] Xerox released an aggregated management forecast, which provided a forecast of earnings only.
[b] Intel released a disaggregated management forecast, which provided a forecast of earnings components.

could affect the processing of product information in print ads.[5] Yi (1990a, 1990b)finds that prior exposure to contextual factors can prime certain product attributes and, in turn, increase the likelihood that consumers interpret product information in terms of these activated attributes. Specifically, Yi (1990a) conducts an experiment in which consumers were given an automobile ad that highlights the car's "large size." Noting that consumers likely have knowledge structures of a car's "large size" that depict the attribute either as a favorable input for car safety or an unfavorable input for fuel efficiency, Yi (1990a) argues that the context surrounding the ad can affect which concept or knowledge structure is activated and used. Specifically, the author predicts that placing the ad within an article about safety primes the "car safety" knowledge structure, making it more likely that consumers interpret the car's large size as a favorable input for car safety. Yi's (1990a) results support his prediction. Consumers provided more favorable judgments about the advertised car when it was located within the car safety article than when it was not.

We contend that investors have at least two general concepts or knowledge structures related to net income. Prior research in accounting suggests that investors sometimes conceptualize net income as one of several important inputs to evaluating a company's prospects (see Elliott et al. 2008), likely obtained from formal education (e.g., see Penman 2007), investing guides, or experience. However, many have suggested that investors often forgo this net income conceptualization and instead conceptualize net income as a sufficient and convenient summary measure of a company's prospects (e.g., see CFA Centre for Financial Market Integrity 2007).[6] The latter conceptualization leads investors to fixate on a bottom-line net income number in subsequent earnings announcements.

We argue that an aggregated forecast isolates net income, priming the conceptualization of net income as a convenient summary measure and, thus, as a substantially more important measure of company prospects relative to its components. By contrast, a disaggregated forecast places net income among a list of other potential inputs in judging the company's prospects. In this case, similar to consumers in Yi's (1990a) study, we propose that a disaggregated forecast places net income in a context that primes or makes salient investors' knowledge structures of net income as one of several similarly relevant inputs in evaluating the company's prospects. Once the investor observes net income in the subsequent earnings announcement, the primed knowledge structure of net income as one of several similarly relevant inputs is activated.[7] Thus, investors' evaluation of net income in assessing the company's prospects is made in terms of the activated knowledge structure. As such, we expect the investment judgments of investors who initially observe a *disaggregated* (an *aggregated*) earnings forecast to display less (more) earnings fixation because they evaluate net income as one of several similarly relevant inputs (as a substantially more

---

[5] Yi (1990a) contemplates both cognitive and affective priming effects on attribute interpretation. Our interest in the effect of forecast display suggests a cognitive priming effect would be more applicable to our study. While Yi (1990a) documents that both cognitive and affective primes can influence attribute interpretation, the author finds that cognitive primes influence attribute interpretation independent of affective primes and that the two types of priming do not interact. Thus, considering affective priming is beyond the scope of our study.

[6] We note that net income can sometimes be a sufficient measure of a company's prospects. Thus, treating net income as a summary measure of a company's prospects may also result from experiencing the past sufficiency of using net income in this manner.

[7] While disclosures other than forecasts may also precede any given earnings announcement, or may occur between a forecast and the related earnings announcement, few, if any, of these other disclosures are likely to cue investors to think of the summary net income figure as one of several similarly important inputs as opposed to the most important input (or at least, not to the same extent as forecasts). In addition, forecasts other than management forecasts, such as analyst forecasts, may also cue investors to conceptualize the summary net income figure differently and could either exacerbate or attenuate our predicted effects depending on the level of aggregation of the analyst forecast, the timing of the analyst forecast disclosure relative to the management disclosure, and investors' desire and ability to access the analyst forecast. We focus on management as opposed to analyst forecasts because we are interested in examining mechanisms that managers can use to reduce investor earnings fixation (other than simply eliminating forecasts altogether).

relevant input) in assessing the company's prospects. Thus, we derive our primary hypothesis as follows:

Hypothesis: After receiving a company's earnings announcement, the level of earnings fixation reflected in investors' investment judgments will be lower (higher) when they initially observe a disaggregated (aggregated) earnings forecast.

III. EXPERIMENTAL METHOD

Participants

Two hundred and one students enrolled in graduate business classes from two large state universities participated in the experiment as reasonably informed investors. Our participants have a reasonable understanding of business, accounting, and finance as, on average, participants had taken 7.61 accounting courses and 3.53 finance courses, and 91 percent of participants had used financial statements to evaluate a company's performance at least one time. In addition, 39 percent of participants stated that they had purchased common stock or debt securities, while 88 percent said that they planned to do so in the next five years.

We use students enrolled in graduate business classes as proxies for reasonably informed investors for two primary reasons. First, as Libby et al. (2002, 802) note, sophisticated participants are in short supply and should only be used when the research question necessitates it. We matched participants with the goals of our experiment. Prior to conducting our experiment, we determined that (1) our task was similar to those described in Elliott et al. (2007) as low in integrative complexity, and (2) a reasonable understanding of financial accounting concepts and basic finance would be necessary and sufficient for participants to respond meaningfully to our experimental materials. Second, prior research that uses tasks with similar characteristics to the tasks in this study (and similar to those classified as low in integrative complexity by Elliott et al. [2007]), finds no substantive differences in the responses of professional analysts and graduate business students (e.g., Barton and Mercer 2005).[8]

Experimental Design and Procedures

To investigate our hypothesis, we use a 2 × 2 between-subjects design. Participants were asked to take on the role of an investor evaluating a hypothetical firm (LearningWare). Participants first received background information that described the company as a developer and seller of various software applications. After examining this background information, participants observed the company's forecast for the year ending December 31, 2007. The forecast was issued on March 30, 2007 and included summary income statements for the years ending December 31, 2005 and 2006 along with the forecast, and suggested a trend in performance consistent with the two previous years. We manipulated the format of the earnings forecast on two levels. Half of the participants received an aggregated earnings forecast (*Aggregated* condition) that aggregated the components of earnings such that only a bottom-line net income measure was provided in the forecast. The remaining participants received a disaggregated forecast (*Disaggregated* condition), in which a forecast was provided for earnings as well as its components (see Appendix A).[9] After

---

[8] In addition, individual investors are an important investor group in their own right. In early 2008, approximately 52.2 million households (45 percent) owned shares of publicly traded stocks, and 16.2 million of these households held stock outside of employer-sponsored plans (Investment Company Institute and the Securities Industry and Financial Markets Association [ICI/SIFMA] 2008). Thus, financial executives and regulators should (and clearly do—http://sec.gov/about/whatwedo.shtml) consider this important group of shareholders, as well as professional investors and analysts, in designing and regulating financial disclosures.

[9] The actual format of the *Aggregated* and *Disaggregated* conditions was adapted from Hirst et al. (2007). Forecasted earnings and its percentage increase were described in text at the top of the forecast press release in the *Aggregated*

examining management's forecast, participants used an 11-point Likert-type scale to indicate the likelihood that they would consider the company as a potential investment and the company's attractiveness as a potential investment (i.e., initial investment judgments are used as covariates in statistical tests).

Upon completing these judgments, participants provided demographic information before continuing the experiment. Participants then received the company's earnings announcement in addition to the same background information and forecast release they had previously seen.[10] The earnings announcements for each condition showed that reported earnings were the same as forecasted earnings.[11] However, the summary income statement showed either higher revenue and research and development expense (*Favorable* condition) or lower revenue and research and development expense (*Unfavorable* condition) than would be expected given the trend of the previous two years (Appendix B).[12]

After examining the company's earnings announcement, participants again indicated the likelihood that they would consider the company as a potential investment and the company's attractiveness as a potential investment (i.e., final investment judgments and primary dependent measure). In addition, participants answered several questions that served as manipulation checks and several questions designed to detect evidence of the process they used in forming their judgments.

Highlights of Design Features

To test a positive (versus negative or null) hypothesis for reducing earnings fixation, we examine when participants discern between two companies whose earnings announcements report identical net income, but whose net income components indicate either more or less favorable future prospects. Manipulating revenue to be less (more) than expected is a clear indicator of unfavorable (favorable) future prospects. However, in order to provide a net income figure that does not change, the *Unfavorable* (*Favorable*) setting requires an income-increasing (-decreasing) event that also can be interpreted as unfavorable (favorable). Therefore, we also choose to manipulate R&D expense; although decreasing (increasing) R&D expense increases (decreases) net income, it can be interpreted as an unfavorable (favorable) indicator.

To reduce the likelihood that differences in information across forecast aggregation levels

---

condition. Forecasts of earnings components (i.e., revenue, gross profit, research and development expense, and earnings) and their percentage changes were described in text at the top of the forecast press release in the *Disaggregated* condition. In addition, as was the case in Hirst et al. (2007), the *Disaggregated* condition included a forecasted summary income statement shown alongside the summary income statements for the prior two years.

10 Asking participants to provide demographic information between receiving the initial forecast and receiving the company's earnings announcement was done to, at least partially, simulate the natural time delay between when a company issues an initial forecast and the subsequent earnings announcement. We acknowledge that, in the natural setting, the time delay would be longer and that this could potentially weaken the results we report below. However, even in the natural setting, investors have access to previously issued forecasts and are often prompted (by management or some other financial intermediary) to reconsider these forecasts in evaluating the realizations reported in the earnings announcement. Prior accounting studies also have allowed participants to access previously issued forecasts (see Libby et al. 2006; Han and Tan 2007).

11 An alternative design choice would have been to manipulate reported earnings and observe whether participants' judgments reflected the difference in reported earnings. In this case, statistically significant judgment differences would indicate earnings fixation and a null result would indicate reduced earnings fixation. Our choice to have identical earnings for all participants but manipulate the underlying components means that evidence for our hypothesis (i.e., reduced earnings fixation) is associated with a significant statistical finding rather than a null finding. Thus, we believe our design choice provides stronger support for our hypothesis.

12 In order to examine boundary conditions for the impact of disaggregating the earnings forecast, we also manipulated the format of the earnings announcement on two levels. Half of the participants received an earnings announcement that emphasized net income by highlighting earnings and its percentage change in the statement heading, while the remaining participants received an earnings announcement that deemphasized net income by highlighting the components of earnings along with earnings itself. We describe this boundary condition test in our "Supplemental Analyses" section.

explain our expected findings, our setting is designed to allow participants across *Forecast Aggregation* conditions to develop the same expectations for the 2007 announced income statement. All participants observe previously announced results in a 2007 forecast that indicate net income and its components for 2006 were roughly 20 percent higher than in 2005. In addition, all participants view a forecast of 2007 net income that exceeds 2006 net income by 20 percent. Thus, the expected increase in net income components in 2007 would likely be 20 percent. Our disaggregated forecast merely makes explicit the likely expectations for the 2007 net income components by forecasting values for 2007 net income components that are 20 percent higher than 2006 amounts. In our discussion of alternative explanations in the "Results" section, we provide evidence that participants had similar expectations across *Forecast Aggregation* conditions.

To reduce the likelihood that the impact of disaggregating the 2007 forecast is explained by management signaling which items are most relevant, our disaggregated forecast provides a forecast for all components of net income. Because management in our setting displays no selectivity over which net income components they forecast, the likelihood that participants perceive the disaggregation as a signal of items that are most important is reduced. In addition to the above design choices, we also collect additional data to address the likelihood of these alternative explanations for our findings in our supplemental analyses discussed at the end of the next section.

IV. RESULTS

Manipulation Checks

To determine whether participants perceived different levels of aggregation across the *Aggregated* and *Disaggregated* conditions, we asked participants to assess the level of detail in the forecast they observed. Using an 11-point Likert-type scale anchored on 1 = Not at All Detailed and 11 = Very Detailed, participants gave mean judgments of forecast detail of 3.97 and 6.07 for the *Aggregated* and *Disaggregated* forecast conditions, respectively. As expected, participants judged the aggregated forecast to have less detail than the disaggregated forecast ($t = 6.78$, $p < 0.01$). In addition, we asked participants to indicate the type of forecast, if any, they observed. Seventy-two percent of participants in the *Aggregated* condition correctly identified that they had received a forecast of net income alone, while 89 percent of participants in the *Disaggregated* condition correctly identified that they had received a forecast of each of the items on the income statement, including net income.[13] Together, these results suggest that participants were sensitive to our manipulation of *Forecast Aggregation* level.

Initial Investment Judgments

Recall that participants provided initial investment judgments after viewing the management forecast (but before viewing the earnings announcement). Even though participants are exposed to one of our manipulated variables before providing their initial investment judgments, we did not expect the level of aggregation in the management forecast to influence participants' initial investment judgments in a predictable manner. Consistent with our expectation, results (not tabulated) reveal that participants' initial investment judgments are not influenced by the level of aggregation in the management forecast to which they were exposed (p-value > 0.50). Thus, in the analyses reported below we simply include participants' initial investment judgments as a covariate.

The Effects of Forecast Disaggregation

Our hypothesis predicts that investors who initially observe a disaggregated management forecast will exhibit less earnings fixation than investors who initially observe an aggregated

---

[13] Our results are inferentially identical if we remove those participants who incorrectly responded to this manipulation check question.

management forecast. We test our hypothesis using participants' judgments of (1) the likelihood of considering the company as a potential investment, and (2) the company's attractiveness as a potential investment as indicated after participants have observed both the forecast and the earnings announcement. A Cronbach coefficient alpha score of 0.92 reveals that the "likelihood of investment" and "investment attractiveness" measures likely measure the same underlying construct. Therefore, we report the results of the hypothesis tests for a single combined investment judgment dependent measure based on the average of the two measures for each participant, adjusted for participants' initial investment judgments (after exposure to only the forecast) as a covariate.[14]

Recall that participants in all conditions observe the same forecasted earnings and the same announced earnings that meet the company's forecast. However, the components that underlie earnings differ. Earnings fixation is thus inferred when participants indicate the same investment judgments when the earnings components differ (i.e., across the *Unfavorable* and *Favorable* conditions). Conversely, we infer a reduction in earnings fixation when participants' investment judgments differ as the earnings components differ.

Results of our experiment support the hypothesis. Table 1, Panel A provides adjusted least square means and standard errors for participants' final investment judgments adjusted for their initial investment judgments as a covariate. Using an 11-point Likert-type scale, anchored on 1 = Not at All Likely/Attractive and 11 = Very Likely/Attractive, participants' final investment judgments overall are more favorable in the *Favorable* condition (mean = 7.30) than in the *Unfavorable* condition (mean = 6.69). An overall ANCOVA (see Panel B of Table 1 and Figure 1), using participants' initial investment judgments as a covariate, reveals a significant interaction between the *Unfavorable/Favorable* and the *Aggregated/Disaggregated* conditions ($F = 3.93$, $p = 0.05$, two-tailed).

Table 1, Panel A indicates the adjusted least square mean investment judgments for participants in the *Aggregated* forecast condition are 7.58 and 7.38 for the *Favorable* and *Unfavorable* conditions, respectively. A simple effects test (see Panel C of Table 1) indicates that the investment judgments of participants in the *Aggregated* forecast condition are no different in the *Favorable* condition (where the components of earnings were more favorable) than in the *Unfavorable* condition (where the components of earnings were less favorable; $t = 0.70$, $p = 0.24$, one-tailed). Turning to participants in the *Disaggregated* forecast condition, the adjusted least square mean investment judgments are 7.02 and 6.00 for the *Favorable* and *Unfavorable* conditions, respectively. In this case, a simple effects test (Table 1, Panel C) indicates a significant difference across the *Favorable* and *Unfavorable* conditions ($t = 3.53$, $p < 0.01$, one-tailed).[15] As predicted, participants in the *Disaggregated* forecast condition distinguish between the company's investment potential when its earnings components were more favorable versus less favorable, suggesting that for these participants earnings fixation is reduced.

---

[14] When analyzing participants' likelihood and attractiveness measures separately, each measure yields the same inferences as the combined investment judgment measure.

[15] *A priori,* the precise pattern of expected cell means is unclear. Our hypothesis specifies that the slope between *Unfavorable* and *Favorable* will differ by the level of aggregation in the management forecast, but it does not suggest a specific pattern of cell means. We expected participants in the *Aggregated/Unfavorable* and *Aggregated/Favorable* conditions to rate investment potential as more favorable than participants in the *Disaggregated/Unfavorable* condition. Specifically, participants in the two former conditions are expected to be more influenced by the realization of the predicted increase in net income, while those in the latter conditions are expected to be more influenced by the failed realizations of predicted increases in revenues and R&D. It is less clear how the investment judgments of participants in the *Disaggregated/Favorable* condition would fall relative to the other conditions. While these participants observed that the company surpassed expectations in reported revenues and spending on R&D, it is somewhat ambiguous as to whether the increased spending on R&D is positive or negative news.

TABLE 1

Descriptive Statistics and Hypothesis Tests

Panel A: Investment Judgment Adjusted Least Squares Means [standard error][a]

	Unfavorable	*Favorable*	Total
Aggregated	7.38	7.58	7.48
	[0.21]	[0.20]	[0.15]
	n = 49	n = 50	n = 99
Disaggregated	6.00	7.02	6.51
	[0.20]	[0.21]	[0.14]
	n = 52	n = 50	n = 102
Column Mean	6.69	7.30	6.99
	[0.14]	[0.15]	[0.15]
	n = 101	n = 100	n = 201

Panel B: Overall ANCOVA

Source	df	Mean Square	F-statistic	p-value[b]
Unfavorable/Favorable	1	18.35	8.80	<0.01
Forecast Aggregation	1	46.97	22.51	<0.01
Unfavorable/Favorable × *Forecast Aggregation*	1	8.19	3.93	0.05
Initial Judgments (COV)	1	335.15	160.61	<0.01

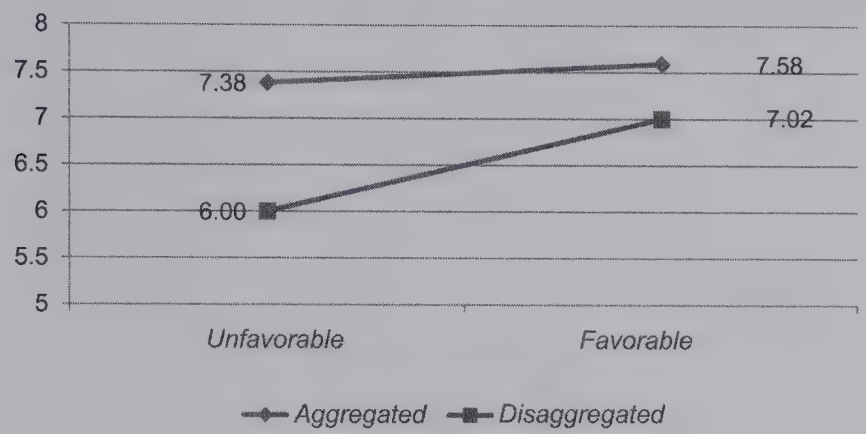
Panel C: Simple Effects Test for Effect of *Forecast Aggregation*

	df	Mean Square	t-statistic	p-value[b]
Unfavorable is not significantly different from *Favorable* when the forecast is aggregated	1	1.02	0.70	0.24
Unfavorable is significantly different from *Favorable* when the forecast is disaggregated	1	25.97	3.53	<0.01

[a] Participants judged their likelihood of considering the company as a potential investment (the attractiveness of the company as an investment) using an 11-point scale anchored on 1 = Not at All Likely (Attractive) and 11 = Very Likely (Attractive). Participants made initial judgments after receiving an *Aggregated* (*Disaggregated*) management forecast of earnings only (earnings and earnings components) and made the same judgments again after receiving an earnings announcement revealing the realization. Participants observed announced earnings that matched forecasted earnings, but whose revenues and R&D expense depicted either *Unfavorable* or *Favorable* performance. The adjusted least square means reported in Panel A were derived by averaging the likelihood and attractiveness responses provided after observing both the forecast and the realization for each participant, adjusted for participants' initial investment judgments (after observing only the forecast) as a covariate.

[b] p-values for the overall ANCOVA are two-tailed; p-values for the simple effects tests are one-tailed given a directional prediction.

FIGURE 1
Effects of Forecast Disaggregation on Investment Judgments



Participants judged their likelihood of considering the company as a potential investment (the attractiveness of the company as an investment) using an 11-point scale anchored on 1 = Not at All Likely (Attractive) and 11 = Very Likely (Attractive). Participants made initial judgments after receiving an *Aggregated* (*Disaggregated*) management forecast of earnings only (earnings and earnings components) and made the same judgments again after receiving an earnings announcement revealing the realization. Participants observed announced earnings that matched forecasted earnings, but whose revenues and R&D expense depicted either *Unfavorable* or *Favorable* performance. The adjusted least square means reported above were derived by averaging the likelihood and attractiveness responses provided after observing both the forecast and the realization for each participant, adjusted for participants' initial investment judgments (after observing only the forecast) as a covariate.

Evidence for Our Theory

Our theory suggests that a disaggregated forecast's ability to reduce earnings fixation is associated with investors who observe a disaggregated forecast interpreting net income to be one of several similarly important inputs in evaluating a company's prospects more than investors who observe an aggregated forecast. To examine this theory, we asked participants to allocate 100 points among the different components of the summary income statement, based on how important they deemed each item for evaluating the company as a potential investment. Fewer (more) points allocated to net income (net income components) would be consistent with participants viewing net income as one of several evaluation inputs rather than as a sufficient summary measure of the company's prospects.

The mean points allocated to net income are 26.30 and 22.82 for participants in the *Aggregated* and *Disaggregated* forecast conditions, respectively (see Table 2, Panel A). Results of an ANOVA test (Table 2, Panel B) reveal a main effect for forecast aggregation ($F = 3.44$, $p = 0.03$, one-tailed), indicating that participants in the *Disaggregated* condition allocate fewer points to net income than participants in the *Aggregated* condition. The result supports our theory about why disaggregated forecasts reduce earnings fixation.[16]

[16] Although an examination of the means reported in Panel A of Table 2 reveals that the points allocated to net income are directionally higher under the *Disaggregated/Favorable* condition than the *Disaggregated/Unfavorable* condition, as revealed in Panel B of Table 2, there is not a significant *Favorable/Unfavorable* by *Forecast Aggregation* interaction ($p = 0.30$).

TABLE 2

Importance of Net Income to Participant Judgments

Panel A: Mean [standard deviation of] Importance of Net Income[a]

	Unfavorable	*Favorable*	**Total**
Aggregated	26.43	26.18	26.30
	[13.19]	[14.59]	[13.84]
	n = 49	n = 50	n = 99
Disaggregated	21.10	24.65	22.82
	[11.43]	[12.92]	[12.25]
	n = 52	n = 49	n = 101
Column Mean	23.68	25.42	24.55
	[12.54]	[13.74]	[13.15]
	n = 101	n = 99	n = 200

Panel B: Overall ANOVA

Source	**df**	**Mean Square**	**F-statistic**	**p-value[b]**
Unfavorable/Favorable	1	136.73	0.80	0.37
Forecast Aggregation	1	587.79	3.44	0.03
Unfavorable/Favorable × *Forecast Aggregation*	1	180.91	1.06	0.30

Panel C: Mean [standard deviation of] Ratio of Points Assigned to Net Income by the Points Assigned to the Most Important Non-Net Income Item

	Unfavorable	*Favorable*	**Total**
Aggregated	1.19	1.30	1.25
	[0.76]	[1.35]	[1.09]
	n = 49	n = 50	n = 99
Disaggregated	0.91	1.16	1.03
	[0.82]	[0.91]	[0.87]
	n = 52	n = 49	n = 101
Column Mean	1.05	1.23	24.55
	[0.80]	[1.15]	[13.15]
	n = 101	n = 99	n = 200

(continued on next page)

Panel D: Overall ANOVA

Source	df	Mean Square	F-statistic	p-value[b]
Unfavorable/Favorable	1	1.61	1.65	0.20
Forecast Aggregation	1	2.28	2.34	0.06
Unfavorable/Favorable × *Forecast Aggregation*	1	0.25	0.26	0.61

[a] Participants allocated 100 points among the components of a hypothetical company's summary income statement (i.e., revenue, cost of sales, gross margin, research and development expense, selling, general, and administrative expenses, and net income) based on how important they believed each metric was for evaluating the company as a potential investment. The dependent measure is mean points allocated to net income. Greater earnings fixation is supported by greater points being allocated to net income. Participants received an *Aggregated* (*Disaggregated*) management forecast of earnings only (earnings and earnings components) and an earnings announcement. In addition, participants observed announced earnings that matched forecasted earnings, but whose revenues and R&D expense either *Unfavorable* or *Favorable* performance.

[b] All p-values shown are two-tailed except for the *Forecast Aggregation* main effect whose p-value is one-tailed, given our directional prediction.

To supplement this evidence, we examine the spread of points allocated across net income and its components. If participants in the *Disaggregated* (*Aggregated*) forecast condition tend to consider net income more (less) as one of several inputs to evaluating the company, we expect the points they assign to the summary net income measure and to the most important net income component to be more (less) similar. Therefore, in an additional analysis, we calculate a ratio for each participant that divides the points assigned to the summary net income measure by the points assigned to the most important net income component (the component with the greatest number of assigned points). Our theory is supported if the calculated ratio is significantly greater in the *Aggregated* forecast condition than in the *Disaggregated* condition. The means of the calculated ratio are 1.25 and 1.03 (Table 2, Panel C) for the *Aggregated* and *Disaggregated* conditions, respectively, and are marginally different ($F = 2.34$, $p = 0.06$, one-tailed, Table 2, Panel D). These findings suggest that participants in the *Aggregated* condition assign 25 percent more points to the summary net income measure than the points they assign to the most important net income component, whereas participants in the *Disaggregated* condition assign a nearly equal number of points to net income and the most important net income component. Thus, participants in the *Disaggregated* condition appear to be more likely than participants in the *Aggregated* condition to consider net income as one of several similarly important inputs, further supporting our theory.[17]

Benchmark Framework as an Alternative Theoretical Explanation

Prior research in accounting has considered a benchmark framework when examining the impact of management forecasts on investor judgments (e.g., Krische 2005; Libby et al. 2006; Han and Tan 2007). An assumption underlying the benchmark framework as applied in investment settings is that investors' judgments might be sensitive to the number of benchmarks missed, met, or exceeded such that when a greater number of benchmarks are met (missed), investors perceive the company as a more (less) favorable potential investment (Han and Tan 2007). Because forecasts represent benchmarks of company performance, a benchmark framework represents an attractive approach to deriving predictions in our setting as an alternative to our theoretical framework.

In applying a benchmark framework to our experimental setting, one could plausibly reason that disaggregated forecasts highlight different benchmarks than aggregated forecasts. For example, participants in the *Disaggregated* forecast condition could use the forecasted line item amounts as benchmarks, while those in the *Aggregated* forecast condition could use prior year line item amounts as benchmarks because they observe no explicit forecasts to serve as benchmarks. More specifically, comparing the net number of benchmarks exceeded when prior year line item amounts are used as benchmarks to the net number of benchmarks exceeded when the forecasts of line items are used as benchmarks can yield a pattern of investment judgments similar to our findings.

Our experimental design does not allow us to discern clearly whether participants in our setting adopted a processing approach more consistent with our theoretical framework or a benchmark framework. Improving our understanding of the processing approach investors adopt when presented with forecasted information (and subsequent realizations) is important given the increasing frequency and importance of forecasts. Thus, parsing out these two theoretical perspectives represents a promising avenue for future research.

---

[17] In addition, the proportion of participants who assigned their greatest number of points to Net Income was greater in the *Aggregated* condition (42 percent) than in the *Disaggregated* condition (33 percent). A one-sided Fisher Exact test (not tabulated) indicates that the difference in proportion is marginally significant ($p = 0.09$).

V. SUPPLEMENTAL ANALYSES

In this section, we examine whether the earnings-fixation-reducing effect of a disaggregated forecast is bounded by characteristics of the earnings announcement. We also provide evidence that our results are not fully explained by three plausible alternative explanations, namely, (1) perceived differences in information availability, (2) a pure management signal of the importance of certain financial statement items, or (3) perceived differences in forecast credibility across the level of forecast aggregation.

The Boundary Effects of Emphasis in Earnings Announcements

Although our study focuses on the ability of forecast disaggregation to reduce earnings fixation in investors' judgments, we include an examination of earnings announcement characteristics to examine boundary conditions for our hypothesis. Specifically, we consider whether an earnings announcement that emphasizes a bottom-line net income number can hinder the ability of a disaggregated forecast to reduce investors' earnings fixation. Investment professionals suggest that the extent to which a report emphasizes net income can affect investors' tendency to fixate on net income. For example, financial analysts assert that the power of a summary earnings number in proposing to eliminate any summary earnings from the income statement, in part, reduces "the focus on a single, arbitrary performance indicator, accounting net income" (CFA Centre for Financial Market Integrity 2007, 30).

We include an *Earnings Emphasis* manipulation in our experiment in which the earnings announcement either emphasizes net income by highlighting earnings and its percentage change in the press release heading, or deemphasizing net income by including in the press release heading the components of net income together with net income itself, and the respective percentage changes from the prior period.[18] We then expand the ANCOVA model (untabulated) used to test our hypothesis to include all main and interactive effects of *Earnings Emphasis* on investors' judgments. Results indicate that whether the earnings announcement emphasizes or deemphasizes earnings has no impact on the ability of disaggregated forecasts to reduce earnings fixation. Specifically, an expanded ANCOVA (untabulated) reveals no three-way interaction between *Forecast Aggregation*, *Earnings Emphasis*, and *Unfavorable/Favorable* ($F = 0.26$, $p = 0.61$, two-tailed). Indeed, in observing the adjusted least square mean investor judgments, the disaggregated forecast appears to reduce earnings fixation when the earnings announcement emphasizes earnings (*Unfavorable* = 6.24 versus *Favorable* = 7.25) by nearly an identical amount as when the announcement deemphasizes earnings (*Unfavorable* = 5.76 versus *Favorable* = 6.76). Moreover, we find no significant interaction between the *Unfavorable/Favorable* and the *Earnings Emphasis* conditions ($F = 0.22$, $p = 0.64$, two-tailed) or the *Forecast Aggregation* and *Earnings Emphasis* conditions ($F = 1.07$, $p = 0.30$, two-tailed). Thus, earnings emphasis in the announcement does not appear to affect the reduction in fixation provided by disaggregating an earnings forecast.[19]

Differences in Available Information

We now consider the possibility that the effect of forecast disaggregation on reducing earnings fixation merely reflects a difference in the forecast information available to participants. As is

[18] Using a Likert-type scale anchored on 1 = Not at all detailed and 11 = Very detailed, participants judged the detail of the emphasized announcement (5.44) and the deemphasized announcement (6.10) to be significantly different ($t = 2.21$, $p = 0.03$).

[19] We had 24 participants take part in an experiment that was equal to our main experiment with the significant exception that no forecasts were given. Thus, this was a direct test of earnings emphasis alone. All 24 participants were placed in the *Unfavorable* condition and were randomly assigned to earnings announcements that either emphasized or deemphasized earnings. We found no significant effect for *Earnings Emphasis* in participant judgments ($F = 1.13$, $p = 0.30$, with initial judgments as the covariate) or points allocated to net income ($t = 0.86$, $p = 0.40$). Thus, in our setting, it does not appear that *Earnings Emphasis* alone eliminates (or even significantly reduces) earnings fixation.

inherently the case in natural settings (and following Hirst et al. [2007]), disaggregated forecasts in our experiment specify a greater number of items than do aggregated forecasts. Thus, one could argue that participants who observed aggregated forecasts ignored the net income components because the lack of forecast information made it impossible to assess the components as positive or negative news. This is likely not the case, however. First, if participants in the *Aggregated* condition were unable to evaluate the net income components, then we would expect some of them to assign approximately zero points to the net income components. The results, in contrast, show that all participants assigned some points to the net income components. Second, we chose the values used in the disaggregated forecast to represent an extension of the linear trend of the 2005 and 2006 income statement amounts. This should allow participants in the *Aggregated* condition to derive an expectation for the components of net income similar to the expectations derived by those in the *Disaggregated* condition.[20] Additional data below suggest that participants in both aggregation levels derive similar expectations.

An investor's surprise following a company announcement represents the extent to which reported amounts deviate from the investor's expectations. Therefore, if participants in the *Aggregated* condition did not form expectations of revenue, research and development, and net income (because they received no forecasts for earnings components), then we would expect them to be less surprised at the amounts reported for these items in the earnings announcement than participants in the *Disaggregated* condition (who received forecasts of the components). To examine this proposition, we ask participants the following question with regard to revenue, research and development, and net income: "How surprised were you by the amount of Revenues/R&D/Net Income reported for 2007?" Participants answered on a Likert-type scale anchored on 1 = "Not at All Surprised" and 11 = "Very Surprised." The mean surprise for announced net income is 5.52 and 5.11 for those in the *Aggregated* and *Disaggregated* forecast conditions, respectively. The mean surprise for announced revenue is 5.47 and 5.82 for those in the *Aggregated* and *Disaggregated* forecast conditions, respectively. And the mean surprise for announced R&D is 6.24 and 6.52 for those in the *Aggregated* and *Disaggregated* forecast conditions, respectively. An ANOVA (not tabulated) reveals no significant main effect for *Forecast Aggregation* (all p-values > 0.30), nor any significant interaction effects (all p-values > 0.49) for any of these dependent measures.[21] Thus, our findings do not seem to be the result of participants who received an aggregated forecast not developing expectations for the components of net income.[22]

Pure Management Signal of Importance of Certain Financial Statement Items

We next consider whether our observed reduction in fixation is driven purely by participants' perception that the level of forecast aggregation is a signal from management regarding the importance of certain financial statement items. In natural settings, companies can be selective about what items to include in the forecast (see Exhibit 1). In such cases, it is reasonable for

[20] Our design choice of a linear trend was made in an attempt to maximize internal validity. Of course, in the natural setting, forecasts of income statement line items do not always match the prior linear trends of those line items. However, we see no reason why our inferences would not extend to such a setting.

[21] Importantly, the level of surprise expressed by participants in the *Unfavorable* condition significantly differed from that expressed by participants in the *Favorable* condition for revenues ($p = 0.02$), R&D ($p < 0.01$), and net income ($p = 0.07$). This indicates that our null result for aggregation was not due to our measure for surprise being inadequately sensitive to detect differences in participants' levels of surprise.

[22] To confirm that participants across aggregation levels attended to the amounts reported in the earnings announcement, we asked participants to recall the amounts reported for 2008 revenue, R&D Expense, and net income. We also asked them to recall how much 2007 amounts for these line items changed in 2008. There were no differences across *Forecast Aggregation* conditions for any of these recall measures (all p-values > 0.10).

investors to conclude that managers might use a disaggregated forecast to highlight those items that are particularly relevant and exclude items that are less relevant. However, our experimental design makes such signaling less likely; recall that our disaggregated forecast includes *all* items reported in the subsequent earnings announcement. The apparent lack of management selectivity should reduce the likelihood that participants perceived the disaggregation as a signal of the particular items that were most important.

Evidence from our data also runs counter to a signaling explanation for our findings. Recall that, in addition to manipulating the aggregation level of forecasts, we also manipulated whether a subsequent earnings announcement emphasized net income. Emphasizing net income or net income components in an announcement would suggest a similar management signal to any signal ascribed to the aggregation level of a forecast. However, as noted previously, we find no evidence that emphasizing earnings in an earnings announcement reduces earnings fixation. Thus, a strict signaling explanation for our overall findings is unlikely.

Forecast Credibility

Hirst et al. (2007) find that forecast aggregation affects investors' judgments of forecast credibility. Therefore, we also consider whether our results can be explained by differences in forecast aggregation causing differences in perceived forecast credibility.[23] Similar to Hirst et al. (2007) we ask participants to judge the credibility of the forecast provided. Participants provided responses on a Likert-type scale with end points, 1 = Not at All Credible and 11 = Very Credible. It should be noted that Hirst et al. (2007) elicit participants' judgments of forecast credibility directly after observing a management forecast (and the participants never observe a subsequent earnings announcement). In contrast, we do not elicit judgments of forecast credibility until after participants observe both the management forecast and the subsequent earnings announcement (so as not to explicitly cue participants to think about forecast credibility before viewing the subsequent earnings announcement and making their final investment judgments).

Our participants' mean ratings of forecast credibility are higher when they observe a disaggregated forecast (mean = 7.15) as compared to an aggregated forecast (mean = 7.00), directionally consistent with the results of Hirst et al. (2007). However, an ANOVA (not tabulated) reveals no main or interactive effects when forecast credibility is the dependent measure and our three independent variables and all of their interactions are included. Thus, it does not appear that differences in participants' judgments of forecast credibility are driving our results. In addition, it is not surprising that our results for the effect of forecast aggregation on participants' judgments of forecast credibility are not as strong as those reported in Hirst et al. (2007), as our participants viewed a realization of the forecast before evaluating forecast credibility.

---

[23] In addition to the credibility question described below, we asked related questions about management's trustworthiness, competence, and credibility; questions about management's opportunities, motives, and incentives to manage earnings; questions about how constrained management was to manage earnings; and a question about the quality of reported net income. No significant differences across aggregation levels were noted for any of the measures (all p-values > 0.15, two-tailed). However, participants' responses did differ across the *Unfavorable* and *Favorable* conditions, as expected. Specifically, participants in the *Unfavorable* condition who observed a cut in R&D (as opposed to an increase in R&D) rated management as less trustworthy ($p = 0.08$, one-tailed), less forthcoming ($p = 0.01$, one-tailed) and less competent ($p = 0.01$, one-tailed) than those in the *Favorable* condition. In addition, participants in the *Unfavorable* condition rated the quality of net income lower than those in the *Favorable* condition ($p = 0.06$, one-tailed). These latter results suggest that participants did attend to management characteristics and the related quality of net income associated with cutting R&D (as opposed to increasing R&D) to arrive at a forecasted realization; however, our participants did not believe that these characteristics differed by the level of forecast aggregation. Finally, it is unclear, *ex ante*, if participants' responses to the questions about management's opportunities, motives, and incentives to manage earnings, and how constrained management was to manage earnings would differ across the *Unfavorable* and *Favorable* conditions. In fact, no significant differences were observed (all p-values > 0.10, two-tailed).

VI. CONCLUSION

This study examines a mechanism intended to reduce investors' susceptibility to earnings fixation (i.e., excessive reliance on accounting earnings as the primary input in investment judgments, without fully considering other information that may be relevant to evaluating the company's investment prospects). We find that investors are less susceptible to earnings fixation when they initially observe a disaggregated management forecast (which forecasts earnings and its components) than when they observe an aggregated forecast (which forecasts only earnings).

Specifically, participants' investment judgments reflect less earnings fixation when they initially observe a disaggregated management earnings forecast. This reduction in earnings fixation appears to be associated with investors interpreting the summary net income figure as one of several similarly important evaluation inputs rather than a substantially more important input (relative to its components). We provide additional evidence that our results cannot be fully explained by several plausible alternative explanations: (1) perceived differences in information availability across levels of aggregation, (2) perceived management signals of the importance of certain financial statement items, or (3) perceived differences in forecast credibility across levels of aggregation. Finally, we provide evidence that suggests our results are not bounded by the level of emphasis on net income in the subsequent earnings announcement.

As with all experimental research, our study is not without limitations. Participants in our study did not have access to all the information they might have available in a natural setting. Moreover, a limitation specific to our study is that an aggregated (disaggregated) forecast, by definition, provides investors with a single (multiple) benchmark(s). Thus, the level of forecast aggregation, in our experiment as well as in the real world, is inexorably linked with the number of benchmarks in the forecast, forming a natural environmental confound. In addition, while we report evidence that a reduction in earnings fixation results from a disaggregated forecast activating an investor-held knowledge structure, we cannot determine whether this activation occurs prior to or subsequent to investors observing the earnings announcement. Finally, because managers may strategically issue a disaggregated forecast to influence investor perceptions of firm value, providing disaggregated forecasts may not always lead investors to optimal judgments.

Despite these limitations, our results provide important insights regarding a potential mechanism to reduce investors' susceptibility to earnings fixation—an institutional phenomenon that can be undesirable to managers and investors. In addition, our study extends the disclosure literature in accounting by providing evidence that a forecast (and its characteristics) can influence how investors interpret subsequently reported accounting information. Our findings should be of interest to managers, who have incentives to reduce pressures for myopic behavior, investors, whose investment returns may be adversely affected by such myopic behavior, and regulators, who have expressed concern about investors' susceptibility to earnings fixation and the resulting impact on manager behavior.

APPENDIX A
DISAGGREGATED MANAGEMENT FORECAST[24]

LearningWare Provides Forecast for 2007

Tallahassee, FL—LearningWare today provided a forecast for the full year, which ends on December 31, 2007. LearningWare expects revenues to increase 20 percent to $43.2 million. The gross profit margin is projected to remain at 49 percent. Research and development expenses are

[24] The forecast shown is for the *Disaggregated* condition. Participants in the *Aggregated* condition received only the last sentence of the text and received only the audited 2005 and 2006 summary income statements.

expected to increase 19 percent to $6.9 million. Selling, general, and administrative expenses are expected to increase 20 percent to $4.9 million. LearningWare expects 2007 net income to increase 22 percent to $6.1 million.

Income Statement (all figures in millions)
Fiscal Year ending December 31

	Audited		**Unaudited**
	2005	**2006**	**2007 *(Forecast)***
Revenue	$30.0	$36.0	$43.2
Cost of Sales	15.3	18.4	22.0
Gross Profit	14.7	17.6	21.2
Research and Development Expense	4.9	5.8	6.9
Selling, General, and & Administrative Expense	3.4	4.1	4.9
Earnings Before Income Taxes	6.4	7.7	9.4
Income Tax Expense	2.2	2.7	3.3
Net Income	$4.2	$5.0	$6.1

About LearningWare: LearningWare (http://www.learningware.com) develops and sells various software applications that help schools educate and evaluate students and help businesses assess the skills of their employees. This news release contains forecasted information that is not audited.

APPENDIX B
UNFAVORABLE/FAVORABLE ANNOUNCED INCOME STATEMENTS[25]

Panel A: Unfavorable Condition

	Audited		
	2005	**2006**	**2007**
Revenue	$30.0	$36.0	$39.6
Cost of Sales	15.3	18.4	20.2
Gross Profit	14.7	17.6	19.4
Research and Development Expense	4.9	5.8	5.1
Selling, General, and Administrative Expense	3.4	4.1	4.9
Earnings Before Income Taxes	6.4	7.7	9.4
Income Tax Expense	2.2	2.7	3.3
Net Income	$4.2	$5.0	$6.1

(continued on next page)

[25] Panels A and B indicate the summary income statements participants observed when in the *Unfavorable* and *Favorable* announcement conditions, respectively. Amounts highlighted (were not highlighted in actual experiment) indicate the fundamental differences between the two conditions although net income is identical. Earnings fixation is reduced (enhanced) to the extent that participants' investment judgments discern (do not discern) between the *Unfavorable* and *Favorable* income statements.

Panel B: Favorable Condition

	Audited		
	2005	**2006**	**2007**
Revenue	$30.0	$36.0	$46.8
Cost of Sales	15.3	18.4	23.9
Gross Profit	14.7	17.6	22.9
Research and Development Expense	4.9	5.8	8.6
Selling, General, and Administrative Expense	3.4	4.1	4.9
Earnings Before Income Taxes	6.4	7.7	9.4
Income Tax Expense	2.2	2.7	3.3
Net Income	$4.2	$5.0	$6.1

REFERENCES

Ajinkya, B. B., and M. J. Gift. 1984. Corporate managers' earnings forecasts and symmetrical adjustments of market expectations. *Journal of Accounting Research* 22 (2): 425–444.

Barton, J., and M. Mercer. 2005. To blame or not to blame: Analysts' reactions to external explanations for poor financial performance. *Journal of Accounting and Economics* 39 (3): 509–533.

Bowen, R. M., X. Chen, and Q. Cheng. 2008. Analyst coverage and the cost of raising equity capital: Evidence from underpricing of seasoned equity offerings. *Contemporary Accounting Research* 25 (3): 657–699.

Bushee, B. J. 1998. The influence of institutional investors on myopic R&D investment behavior. *The Accounting Review* 73 (3): 305–333.

———. 2001. Do institutional investors prefer near-term earnings over long-run value? *Contemporary Accounting Research* 18 (2): 207–246.

CFA Centre for Financial Market Integrity. 2006. Breaking the short-term cycle. *CFA Institute Centre Publications*: 1–20. Available at: http://www.cfapubs.org/doi/pdf/10.2469/ccb.v2006.n1.4194.

———. 2007. A comprehensive business reporting model: Financial reporting for investors. *CFA Institute Centre Publications*: 1–59. Available at: http://www.cfapubs.org/doi/abs/10.2469/ccb.v2007.n6.4818.

———. 2008. Short-termism survey: Practices and preferences of investment professionals. *CFA Institute Centre Publications*: 1–33. Available at: http://www.cfainstitute.org/Survey/short_termism_survey_report_may_2008.pdf.

Chen, S., D. Matsumoto, and S. Rajgopal. Forthcoming. Is silence golden? An empirical analysis of firms that stop giving quarterly earnings guidance. *Journal of Accounting & Economics*.

Cheng, M., K. R. Subramanyam, and Y. Zhang. 2007. Earnings guidance and managerial myopia. Working paper, University of Southern California.

Cotter, J., A. I. Tuna, and P. D. Wysocki. 2006. Expectations management and beatable targets: How do analysts react to explicit earnings guidance? *Contemporary Accounting Research* 23 (3): 593–624.

Economist, The. 2006. The sound of silence. *The Economist* (April): 29.

Elliott, W. B., F. D. Hodge, J. J. Kennedy, and M. Pronk. 2007. Are M.B.A. students a good proxy for nonprofessional investors? *The Accounting Review* 82 (1): 139–168.

———, ———, and K. E. Jackson. 2008. The association between nonprofessional investors' information choices and their portfolio returns: The importance of investing experience. *Contemporary Accounting Research* 25 (2): 473–498.

Fischhoff, B. 1977. Perceived informativeness of facts. *Journal of Experimental Psychology. Human Perception and Performance* 3 (2): 349–358.

Graham, J. R., C. R. Harvey, and S. Rajgopal. 2005. The economic implications of corporate financial reporting. *Journal of Accounting and Economics* 40 (1–3): 3–73.

Han, J., and H.-T. Tan. 2007. Investors' reactions to management guidance forms: The influence of multiple benchmarks. *The Accounting Review* 82 (2): 521–543.

Herz, R. H. 2006. Testimony of Robert H. Herz, Chairman, Financial Accounting Standards Board before the Capital Markets, Insurance and Government Sponsored Enterprises Subcommittee of the Committee on Financial Services. March 29. Available at: http://www.fasb.org/cs/ContentServer?c=Document_C&pagename=FASB%2FDocument_C%2FDocumentPage&cid=1176156900579.

Higgins, E. T., and G. A. King. 1981. Accessibility of social constructs: Information processing consequences of individual and contextual variability. In *Personality, Cognition, and Social Interaction*, edited by Kihlstrom, J. F., and N. Cantor. Hillsdale, NJ: L. Erlbaum Associates.

Hirst, D. E., L. Koonce, and S. Venkataraman. 2007. How disaggregation enhances the credibility of management earnings forecasts. *Journal of Accounting Research* 45 (4): 811–837.

Houston, J. F., B. I. Lev, and J. Tucker. 2010. To guide or not to guide? Causes and consequences of stopping quarterly earnings guidance. *Contemporary Accounting Research* 27 (1).

Hutton, A. P., and P. Stocken. 2009. Prior forecasting accuracy and investor reaction to management earnings forecasts. Working paper, Boston College and Dartmouth College.

Investment Company Institute and the Securities Industry and Financial Markets Association (ICI/SIFMA). 2008. Equity and bond ownership in America. Available at: http://www.ici.org/pdf/rpt_08_equity_owners.pdf.

Janjigian, V., and M. Ozanian. 2006. Gimmie guidance. *Forbes.com.*

Johnson, M. F., R. Kasznik, and K. K. Nelson. 2001. The impact of securities litigation reform on the disclosure of forward-looking information by high technology firms. *Journal of Accounting Research* 39 (2): 297–327.

Krische, S. D. 2005. Investors' evaluations of strategic prior-period benchmark disclosures in earnings announcements. *The Accounting Review* 80 (1): 243–268.

Lansford, B., B. Lev, and J. W. Tucker. 2010. Causes and consequences of disaggregating earnings guidance. Working paper, Northwestern University, New York University, and University of Florida.

Libby, R., R. Bloomfield, and M. Nelson. 2002. Experimental research in financial accounting. *Accounting, Organizations and Society* 27: 777–812.

———, and H.-T. Tan. 1999. Analysts' reactions to warnings of negative earnings surprises. *Journal of Accounting Research* 37 (2): 415–435.

———, ———, and J. E. Hunton. 2006. Does the form of management's earnings guidance affect analysts' earnings forecasts? *The Accounting Review* 81 (1): 207–225.

McCafferty, J. 2007. Companies are trying to shift investors' focus from short-term results to long-term goals. *CFO Magazine* (May 1).

Murray, D. 1991. Data fixation: Methodological refinements and additional empirical evidence. *Behavioral Research in Accounting* 3: 25–38.

Penman, S. H. 2007. *Financial Statement Analysis and Security Valuation*. New York, NY: McGraw-Hill/Irwin.

Schrand, C. M., and B. R. Walther. 2000. Strategic benchmarks in earnings announcements: The selective disclosure of prior-period earnings components. *The Accounting Review* 75 (2): 151–177.

Skinner, D. J. 1994. Why firms voluntarily disclose bad news. *Journal of Accounting Research* 32 (1): 38–60.

———. 1997. Earnings disclosures and stockholder lawsuits. *Journal of Accounting and Economics* 23 (3): 249–282.

Tan, H.-T., R. Libby, and J. E. Hunton. 2002. Analysts' reaction to earnings preannouncement strategies. *Journal of Accounting Research* 40 (1): 223–246.

Taylor, S. E., and S. T. Fiske. 1979. Salience, attention, and attribution: Top of the head phenomena. In *Advances in Experimental Social Psychology*, edited by Berkowitz, L. New York, NY: Academic Press.

Tversky, A., and D. Kahneman. 1973. Availability: A heuristic for judging frequency and probability. *Cognitive Psychology* 5 (2): 207–232.

———, and ———. 1974. Judgment under uncertainty: Heuristics and biases. *Science* 185 (4157): 1124–1131.

Wyer, R. S., and T. K. Srull. 1981. Category accessibility: Some theoretical and empirical issues concerning

the processing of social stimulus information. In *Social Cognition: The Ontario Symposium*, edited by E. T. Higgins, C. P. Herman, and M. P. Zanna, 161–197. Hillsdale, NJ: Erlbaum.

Yi, Y. 1990a. Cognitive and affective priming of the context for print advertisements. *Journal of Advertising* 19 (2): 40–48.

———. 1990b. The effects of contextual priming in print advertisements. *The Journal of Consumer Research* 17 (2): 215–222.

THE ACCOUNTING REVIEW
Vol. 86, No. 1
2011
pp. 209–236

American Accounting Association
DOI: 10.2308/accr.00000011

Percent Accruals

Nader Hafzalla
Russell Lundholm
University of Michigan

E. Matthew Van Winkle
Voyant Advisors, LLC

ABSTRACT: We document how the effectiveness of an accruals-based trading strategy changes with the benchmark used to identify an extreme accrual. We measure "percent accruals" as accruals scaled by earnings, rather than total assets, and show that this seemingly small change produces a radically different sort of the data. We find that a trading strategy based on percent accruals yields significantly larger annual hedge returns than the traditional accruals measure, and does so mostly by improving the long position in low-accrual stocks. The hedge returns are also significant in all but the lowest quintile of arbitrage risk. We show that percent accruals more effectively select firms where the difference between sophisticated and naïve forecasts are the most extreme. As such, our results are consistent with the earnings fixation hypothesis and are inconsistent with some alternative explanations for the accrual anomaly.

Keywords: *accruals; market inefficiency; abnormal returns.*

Data Availability: *Data used in this study are available from the sources identified in the text.*

I. INTRODUCTION

The accruals anomaly, originally identified by Sloan (1996), is one of the most important findings in accounting research. By purchasing firms with low accruals and selling short firms with high accruals, Sloan (1996) documents a significant hedge return. A large literature has followed exploring the cause of this market inefficiency, but also questioning the size and persistence of the anomaly. In the vast majority of these studies, accruals are measured as earnings less cash from operations, scaled by average total assets. In this study we propose a small change in the definition of accruals that yields surprisingly different results. Instead of scaling

We thank Rachel Hayes, Matthew Magilke, Richard Sloan, Teri Yohn, and the workshop participants at Barclay's Global Investor, University of British Columbia, University of Massachusetts, University of Michigan, The Pennsylvania State University Research Conference, The University of Utah, and Washington University for helpful comments.

Nader Hafzalla died June 29, 2006 while in the fourth year of his Ph.D. program at the University of Michigan. He was a talented scholar and a terrific friend. This work is based on two separate projects that Nader was involved in and is dedicated to his memory.

Editor's note: Accepted by Steven Kachelmeier, with thanks to Dan Dhaliwal for serving as editor on a previous version.

Submitted: February 2008
Accepted: May 2010
Published Online: January 2011

accruals by average total assets, we scale by earnings. We label this definition of accruals as "percent accruals" and argue that it is a more natural interpretation of the original idea that investors are fixated on reported earnings and do not distinguish between accruals and cash flows.

Percent accruals identify an extreme observation as one for which the accrual component makes up a large positive or negative fraction of the total earnings for that year.[1] This measure of accruals focuses on the composition of earnings, distinguishing how much is cash and how much is accrual. We show that this seemingly simple change in the definition of an extreme accrual produces a radically different sort of the data; only 12 percent of the data in the first decile of the traditional accruals measure are in the first decile of our new measure. By simply changing the benchmark for an extreme accrual, we find a hedge return that is much larger than the return based on the traditional definition of accruals, and is particularly successful for the long position.

The percent accruals measure differs most from the traditional accruals measure in the lowest decile—the firms for which a trading strategy takes a long position. The returns to this part of a hedge trading strategy are particularly important because the transaction costs and other limits to arbitrage are significantly lower on a long position than on a short position. We find a 5.53 percent abnormal return to the long position in the first decile of percent operating accruals that is significant at the 0.001 level. This return is four times as large as the 1.27 percent abnormal return in the first decile of the traditional accruals measure, which is not significantly different from zero.[2] Further, the firms in the first decile of percent operating accruals have an average market value of $1,510 million, as compared to only $474 million for the first decile of the traditional operating accruals measure. This suggests that the transaction costs to trading on a percent accruals anomaly are probably much lower than for the traditional accruals measure, and make the evidence for inefficient pricing of percent accruals even more compelling. In addition, we show that the hedge returns based on percent accruals are significant in all but the lowest quintile of arbitrage risk, making the results even harder to reconcile with an efficient market.

Because *percent accruals*—accruals measured as a percent of absolute income—is a new definition of accruals, we examine what types of firms end up in the extreme high and low portfolios of this measure, and why they might experience excess positive or negative returns in the future. The typical firm in the extreme negative percent operating accrual portfolio has large positive cash from operations, but then accrues income back down to something much closer to zero; similarly, to end up in the extreme positive percent operating accrual portfolio, a typical firm has large negative cash from operations, but accrues income up to something near zero. We show that these extreme combinations of cash from operations and accruals are exactly the combinations that produce the most extreme differences between a sophisticated income forecast—one that distinguishes between cash flows and accruals—and a naïve forecast. As a result, the percent accruals measure is arguably superior to the traditional accruals measure at identifying cases for which the earnings fixation hypothesis predicts the largest mispricing.

The percent accruals measure has other advantages. We show that the extreme percent accruals are not disproportionately sensitive to special items, as Dechow and Ge (2006) report for traditional operating accruals, and the hedge returns are equally large in the subsample with or without special items. We also show that the long position is more effective in the subsample of

---

[1] In order to assure that the sign of the accrual is maintained when we sort firms into portfolios, we scale by the absolute value of net income. Because net income can be at or near zero, the result will sometimes be extreme. Note, however, that we only use this variable to sort firms into ten portfolios; beyond placing a firm into the first or tenth portfolio, the extremity of the variable is irrelevant.

[2] As discussed in the next section, comparing the hedge returns from different accruals studies is difficult due to a prevalent look-ahead bias documented in Kraft et al. (2006).

firms with losses than those with gains, unlike the traditional accrual measure that has been shown to have no predictive power for future returns when the sample is limited to loss firms (Dopuch et al. 2009).

Aside from offering new evidence consistent with the accruals anomaly, our study contributes to the growing literature that attempts to explain the cause of the anomaly. The main question in this literature is whether the anomaly is driven by investors' failure to understand the different mean-reversion tendencies of cash flows versus accruals—commonly known as the earnings fixation hypothesis. One inconsistency between prior accrual anomaly results and this hypothesis is that existing studies find very little evidence of mispricing in the low-accrual portfolio.[3] An alternative to the earnings fixation hypothesis is offered by Kothari et al. (2006). They argue that managers of overpriced stocks manipulate earnings up to maintain the overvaluation for as long as possible, the market is fooled by these positive accruals, and this causes the high-accrual portfolio to earn a significant negative return when the accruals reverse and the market corrects itself. In contrast, managers of undervalued firms have no incentive to manipulate earnings down, and consequently there is no predicted undervaluation for the low-accrual portfolio, consistent with the near-zero returns observed in this portfolio. Our results reinstate the significant underpricing of the low-accrual portfolio when accruals are measured relative to earnings instead of total assets. This is inconsistent with the asymmetric prediction of Kothari et al.'s (2006) agency hypothesis. More generally, since neither the earnings fixation hypothesis nor the agency hypothesis is conditioned on exactly how accruals are scaled, our results suggest that a more refined theory is in order.

Another competitor to the earnings fixation hypothesis is offered by Fairfield et al. (2003), who note that when accruals are scaled by assets, the resulting measure is essentially the percentage change in operating assets. If investors fail to understand that investments tend to have diminishing returns, then they may overprice large increases in assets and underprice large decreases in assets. Rather than fixating on earnings, investors simply fail to understand the economics of investment. Our results do not invalidate this interpretation of the past accruals literature or the evidence supporting the "growth" hypothesis, but this interpretation does not apply to our rescaled measure of accruals. Finally, as discussed above, the percent accruals measure produces a better implementation of the trading strategy predicted by the earnings fixation hypothesis than the traditional accruals measure, and it generates superior excess returns, lending further evidence in support of the earnings fixation hypothesis.

Two other studies are related to ours. Dopuch et al. (2009) show that the traditional accrual anomaly holds only for profit firms; the hedge return in the loss firm sample is not statistically different from zero. This finding is due primarily to loss firms with large negative accruals; this portfolio is supposed to have positive excess returns in the accrual anomaly, yet the returns are −0.2 percent in the loss sample. Firm-years with losses comprise roughly one third of the sample, so this is not an inconsequential limitation. In contrast, the return to a hedge strategy based on percent accruals is slightly larger in the loss subsample than in the gain subsample, and is significant in both samples.

Cheng and Thomas (2005) assess 22 different abnormal accruals models and 5 different design choices as a comprehensive investigation into the specifications used to study the accruals anomaly. They find significant variation in the size of the anomaly over the 704 permutations considered. However, all of their accruals models are scaled by some measure of firm size, either total assets or market value, but never by earnings.

---

[3] See Kraft et al. (2006) and Zach (2004) specifically for this topic, as well as Beaver et al. (2007), Beneish and Vargus (2002), D'Avolio et al. (2002), Houge and Loughran (2000), and Lesmond and Wang (2006).

In the next section we describe our new definition of extreme accruals, in Section III we discuss the sample properties of the different accrual measures, and in Section IV we present stock return results based on our new measures. We investigate the cause of the percent accruals' success in Section V and conclude in Section VI.

II. IDENTIFYING EXTREME ACCRUALS

To begin we define the two "traditional" accrual measures found in the literature. The most common definition is:

$$\begin{aligned}\text{Traditional Operating Accruals} = {} & (\text{Net Income} \\ & - \text{Cash from Operations})/\text{Average Total Assets}.\end{aligned}$$

Sloan's (1996) original definition was based on balance sheet changes but, starting with Collins and Hribar (2002), the data from the statement of cash flows have become the preferred definition because they exclude the effects of acquisitions and foreign currency translation adjustments. We use the same variables as in Kraft et al. (2006), calculating traditional operating accruals as the difference between net income (Compustat item 172) and cash from operations (item 308), then dividing by the average of total assets (item 6).

Recently, Richardson et al. (2005) introduced a more comprehensive measure of accruals, specified as:

$$\begin{aligned}\text{Traditional Total Accruals} = {} & [\text{Net Income} \\ & - (\text{Net Dividends and Distributions to/from Equityholders} \\ & + \text{increase in the cash balance})]/\text{Average Total Assets}.\end{aligned}$$

Traditional total accruals are also computed using the cash flow statement, as in Richardson et al. (2005), calculated here as the difference between income before extraordinary items (item 172) less total operating, investing, and financing cash flows (items 308, 311, and 313) plus sales of common stock (item 108) less stock repurchases and dividends (items 115 and 127), divided by the average of total assets (item 6).

Note that both definitions of accruals, and all those we could find in the literature (see Cheng and Thomas 2005), scale by some measure of firm size, in this case total assets. So firms in the first decile of accruals have large negative accruals relative to their size. Another natural benchmark for extremity would be relative to the firm's earnings for the period. By computing accruals as a percent of earnings, the measure focuses on the *composition* of earnings—the relative amounts of cash and accruals. To assure that negative accruals are in the lower portfolios and positive accruals are in higher portfolios, we scale by the absolute value of earnings. That is, our percent accruals measures are:

$$\text{Percent Operating Accruals} = (\text{Net Income} - \text{Cash from Operations})/|\text{Net Income}|$$

and

$$\begin{aligned}\text{Percent Total Accruals} = {} & [\text{Net Income} \\ & - (\text{Net Dividends and Distributions to/from Equityholders} \\ & + \text{increase in the cash balance})]/|\text{Net Income}|.\end{aligned}$$

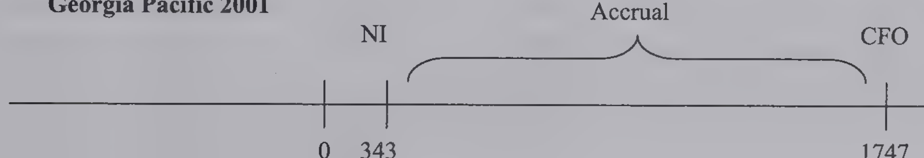
For the percent accruals measures, firms in the first accruals decile have large negative accruals and small positive or small negative earnings; this combination yields a large negative numerator and a small denominator. The cutoff to be in the lowest decile of percent operating

accruals varies by year, ranging from -5.7 to -3.1. This implies that all the firms in this decile have positive cash from operations and their negative accrual pushes their net income back toward zero (if cash from operations was negative, then a negative accrual would only push net income further from zero, making the denominator greater than the numerator and the resulting value greater than -1). Similarly, firms in the highest decile of percent accruals tend to have negative cash from operations and their positive accrual pushes their net income back toward zero. Because the denominator is the absolute value of net income, and net income can be near zero, some values of the percent accruals are extreme.[4] Note, however, that this variable is only used to sort firms into portfolios; extreme values have no more impact on the results than do the more moderate values within the same portfolio.

Consider the examples of Georgia Pacific and Supervision, as shown in Figure 1, which are taken from the lowest decile of percent operating accruals. Both examples illustrate how a large positive cash flow and a large negative accrual combine to yield an extreme negative value of percent operating accruals. These examples also illustrate why the percent operating accruals measure is insensitive to firm size, unlike the traditional operating accruals measure: Georgia Pacific has earnings and cash flows (and unreported total assets) that are orders of magnitude greater than Supervision's numbers, yet they have almost identical percent operating accruals. As we show later, this results in significantly larger firms in the first decile of percent operating accruals than in the first decile of traditional operating accruals. These examples also illustrate why it makes sense to use the absolute value of earnings in the denominator, as one would not want to treat these two very similar examples differently. Finally, these examples illustrate why very poorly performing firms are unlikely to be sorted into the first decile of a percent accruals measure. Firms in the lowest percent operating accruals decile have positive cash flows and

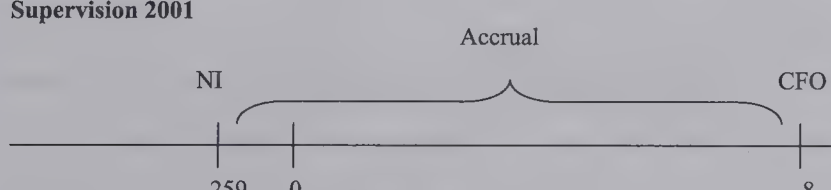
FIGURE 1
Examples from the First Decile of Percent Operating Accruals

Georgia Pacific 2001



Percent Operating Accrual = (343 – 1747)/|343| = -4.09

Supervision 2001



Percent Operating Accrual = (-.259 – .8)/|-.259| = -4.08

---

4 Five observations have zero net income. We sorted these into either decile 1 or decile 10, depending on the sign of the accrual.

earnings near zero, while firms in the lowest traditional operating accruals decile have large losses and large negative cash flows. Consequently, sorting firms by percent accruals measures will spread poor performing observations across the distribution, rather than concentrate them in the lowest decile. We document these empirical regularities in the next section.

III. SAMPLE AND DESCRIPTIVE STATISTICS

We use the merged 2008 Compustat/CRSP database on the WRDS system to collect all firms with available data to construct our accrual measures. Consistent with prior studies, we exclude financial firms. The only data requirements are that the firm has sufficient financial statement data on Compustat to compute traditional operating accruals in the current year, and it has return data available at the portfolio formation date (the first day of the fourth month after the fiscal year-end). The result is a sample of 81,526 firm-years spanning the years 1989–2008. As discussed later, we use the prior year's decile breaks to form portfolios, resulting in 19 years of annual portfolio returns for our returns tests. Table 1 reports descriptive statistics. Because our data requirements are so minimal, the sample is representative of the nonfinancial market as a whole for this time period.

TABLE 1

Descriptive Statistics

Variable	Mean	Median	Standard Deviation	Lower Quartile	Upper Quartile
Percent operating accruals	−1.6860	−0.6753	5.2630	−1.6040	−0.0644
Traditional operating accruals	−0.0670	−0.0517	0.1345	−0.1077	−0.0059
Percent total accruals	−0.8576	0.1737	5.2010	−1.0090	1.0000
Traditional total accruals	−0.0730	0.0084	0.2784	−0.0986	0.0668
Return on Assets	−0.0326	0.0311	0.2235	−0.0503	0.0764
Cash Return on Assets	0.0349	0.0685	0.1765	−0.0062	0.1285
Total Cash Flow Return on Assets	−0.1077	−0.0017	0.2875	−0.0944	0.0341
Market Value of Equity	1,696	138	5,449	30	743
Book Value/Market Value	0.6310	0.4949	0.6239	0.2698	0.8236
Price Per Share	16.61	10.72	17.16	3.88	24.00
Three-Year Sales Growth	0.1076	0.0589	0.4799	−0.1025	0.1869
Proportion of firms with losses	0.3445	1.0000	0.4752	0.0000	1.0000
% of Firms with Special Item <0	0.1988	0.0000	0.3991	0.0000	0.0000
Arbitrage Risk	0.1516	0.1331	0.0840	0.0912	0.1892

The sample period is 1988–2008, consisting of 81,526 firm-years. Traditional operating accruals are defined as Net Income (Compustat item 172) less Cash from Operations (item 308) divided by average total assets (item 6). Traditional total accruals are computed as Net Income less Net Dividends/Distributions (item 115 + item 127 − item 108) and the change in cash (item 308 + item 311 + item 313) divided by average total assets. Percent Operating Accruals and Percent Total Accruals have the same numerators as Operating Accruals and Total Accruals, respectively, but the denominator for each is the absolute value of Net Income. Return on Assets is Net Income divided by average total assets. Cash Return on Assets is cash from operations divided by average total assets. Total Cash Flow Return on Assets are Net Dividends/Distributions plus the change in cash divided by average total assets. Market Value of Equity is measured at the fiscal year-end (item 25 times item 199). Price (item 199) and Book Value (item 60) are measured at the fiscal year-end. Three-year Sales growth is the average annual sales growth in the current year and prior two years. Special items (item 17) are coded as 0 if missing. Arbitrage Risk is measured as the standard deviation of the residuals from a regression of firm-specific returns on the CRSP equally weighted index for up to 48 months preceding the month of return portfolio formation. Amounts are Winsorized at 1 percent and 99 percent for this table and Table 2 only.

Table 2 gives the distribution of the four accruals measures as well as the distributions of other firm characteristics across deciles. The first row in Panel A shows that to be in the lowest decile of percent operating accruals in the average year, the negative accrual needs to be at least 4.04 times earnings; to be in the highest decile, the positive accrual needs to be at least 0.82 times earnings. Comparing Panel A to the traditional operating accruals in Panel B shows that the types of firms that are sorted into the extreme deciles are quite different. The mean market value in decile 1 of percent operating accruals is \$1,510 million versus \$474 million for traditional operating accruals. The firms in decile 1 of percent operating accruals are also performing much better, with positive average return on assets and cash return on assets; in contrast, the firms in the first decile of traditional operating accruals are among the worst performers in the sample with the lowest mean return on assets and cash return on assets. At the other extreme, the firm size and performance in the tenth decile are very similar in Panels A and B. As we show in the next table, this is because the firms overlap significantly at this end of the two different accrual sorts. Finally, Panel A shows that the percent operating accruals measure sorts on cash flows to some degree. Because Desai et al. (2004) document a trading strategy based on cash flows and contrast it with an accruals trading strategy, in the following sections we examine the overlap between the percent accruals measures and cash flow measures. The remaining descriptive statistics will be discussed in Section V as part of the exploration into why percent accruals are successful in identifying misvalued stocks.

Panels C and D of Table 2 give the distributions for the total accruals measures. Comparing the firms across the distributions of these two measures yields similar conclusions to the findings for the operating accrual measures, with one exception: sorting on percent total accruals does not sort on the cash return on assets.

Table 3 gives evidence on the overlap between a sort on percent accruals and a sort on traditional accruals or cash from operations. Generally, the overlap between percent operating accruals and traditional operating accruals is small in the low deciles but higher in the high deciles. Only 11.78 percent of the firms in the first decile of percent operating accruals are in the first decile of traditional operating accruals. This increases to 72.95 percent overlap in the tenth decile. A more general way to document the overlap between the two sorts is to give the average decile rank of one measure in each decile of the other measure. The decile rank of traditional operating accruals is 3.49 in the first decile of percent operating accruals (versus 1 if the two sorts were identical), and increases only marginally until the sixth decile where the average rank is 4.69. The average rank in the tenth decile is 9.70.

The sort on percent operating accruals produces a weak inverse sort on cash flows, as discussed earlier. The mean decile rank of cash from operations is 7.07 in the lowest percent operating accrual decile and is 2.44 in the highest decile (versus 10 and 1, respectively, if the two sorts were identical). The match between the two sorts is weak, however, with the average cash flow rank remaining above five until the ninth decile of percent operating accruals. In Section V we show that the returns to a percent operating accruals trading strategy are not driven by its correlation with cash from operations.

Table 3, Panel B shows the overlap between percent and traditional total accruals. For total accruals, the overlap is a bit stronger for low accruals and a bit weaker for high accruals. There is a 23.17 percent overlap in the first decile but only a 46.27 percent overlap in the tenth decile. The relevant cash flow comparison for total accruals is the net dividends/distributions plus the change in cash (i.e., total cash flows). Column three of Panel B shows that percent total accruals is picking out the extremes for total cash flows, with an average rank of 8.60 and 2.03 in the first and tenth deciles, respectively, although it does not appear to sort well through the middle of the distribution.

TABLE 2

Descriptive Statistics by Deciles of Various Accrual Measures

Panel A: Deciles Sorted by Percent Operating Accrual

	1	2	3	4	5	6	7	8	9	10
Maximum Percent Operating Accrual	−4.04	−2.04	−1.34	−0.96	−0.67	−0.42	−0.18	0.14	0.82	
Mean Market Value	1,510	1,676	1,984	1,912	1,867	2,121	2,269	1,823	1,248	515
Mean Return on Assets	0.003	0.002	−0.014	−0.065	−0.091	−0.085	−0.093	−0.055	0.031	0.044
Mean Cash Return on Assets	0.118	0.117	0.108	0.084	0.051	0.017	−0.027	−0.034	−0.010	−0.079
Mean of Three-Year Sales Growth	0.082	0.091	0.084	0.088	0.103	0.119	0.130	0.135	0.129	0.116
Mean Book-to-Market	0.882	0.794	0.715	0.650	0.580	0.514	0.471	0.474	0.544	0.686
Percent of Firms with Positive Earnings	0.629	0.632	0.621	0.573	0.586	0.623	0.627	0.673	0.789	0.809
Percent of Firms with Special Item <0	0.228	0.212	0.227	0.267	0.262	0.194	0.157	0.114	0.089	0.090
Arbitrage Risk	0.139	0.138	0.135	0.148	0.156	0.155	0.159	0.159	0.158	0.170
Price per Share	13.92	15.66	17.41	16.70	17.38	18.93	19.01	18.18	16.83	12.03

Panel B: Deciles Sorted by Traditional Operating Accrual

	1	2	3	4	5	6	7	8	9	10
Maximum Regular Operating Accrual	−0.20	−0.13	−0.09	−0.07	−0.05	−0.03	−0.02	0.01	0.06	
Mean Market Value	474	1,265	1,852	2,294	2,531	2,573	2,244	1,888	1,218	591
Mean Return on Assets	−0.383	−0.088	−0.024	0.004	0.014	0.023	0.024	0.032	0.031	0.047
Mean Cash Return on Assets	−0.032	0.070	0.085	0.085	0.075	0.066	0.050	0.037	0.002	−0.092
Mean of Three-Year Sales Growth	0.128	0.108	0.106	0.101	0.090	0.098	0.097	0.105	0.109	0.134
Mean Book-to-Market	0.501	0.663	0.667	0.672	0.662	0.661	0.654	0.656	0.616	0.557
Percent of Firms with Positive Earnings	0.132	0.415	0.595	0.691	0.744	0.787	0.796	0.800	0.799	0.801
Percent of Firms with Special Item <0	0.550	0.325	0.230	0.163	0.127	0.099	0.090	0.085	0.090	0.084
Arbitrage Risk	0.217	0.171	0.148	0.134	0.128	0.125	0.128	0.136	0.151	0.179
Price per Share	7.14	12.56	16.45	19.21	20.56	21.19	20.72	18.72	16.27	13.27

(continued on next page)

Panel C: Deciles Sorted by Percent Total Accrual

	1	2	3	4	5	6	7	8	9	10
Maximum Percent Total Accrual	−1.44	−0.89	−0.44	0.01	0.35	0.66	0.93	1.29	2.44	
Mean Market Value	1,247	770	1,067	1,946	2,839	3,103	2,369	1,446	1,092	779
Return on Assets	−0.023	−0.181	−0.201	−0.110	0.003	0.058	0.071	0.071	0.054	0.024
Cash Return on Assets	0.071	0.001	−0.040	−0.016	0.057	0.094	0.093	0.082	0.058	0.021
Total Cash Flow Return on Assets	0.099	0.024	−0.042	−0.060	−0.011	0.010	−0.002	−0.024	−0.076	−0.177
Mean Three-Year Sales Growth	0.072	0.074	0.116	0.096	0.102	0.109	0.111	0.107	0.111	0.108
Mean Book-to-Market	0.814	0.765	0.643	0.585	0.569	0.562	0.582	0.604	0.635	0.685
Percent of firms with positive earnings	0.462	0.172	0.294	0.563	0.787	0.894	0.940	0.939	0.891	0.777

Panel D: Deciles Sorted by Traditional Total Accrual

	1	2	3	4	5	6	7	8	9	10
Maximum Regular Total Accrual	−0.18	−0.07	−0.03	0.00	0.02	0.04	0.06	0.10	0.16	
Mean Market Value	402	919	1,465	2,148	2,361	2,534	2,342	2,008	1,533	959
Return on Assets	−0.399	−0.102	−0.023	0.008	0.025	0.037	0.048	0.059	0.070	0.050
Cash Return on Assets	−0.124	0.015	0.055	0.067	0.073	0.077	0.081	0.082	0.076	0.022
Total Cash Flow Return on Assets	−0.034	0.011	0.023	0.021	0.016	0.009	−0.001	−0.016	−0.054	−0.230
Mean Three-Year Sales Growth	0.124	0.084	0.077	0.076	0.081	0.085	0.095	0.110	0.126	0.148
Mean Book-to-Market	0.474	0.756	0.793	0.809	0.758	0.707	0.638	0.575	0.504	0.439
Percent of firms with positive earnings	0.096	0.269	0.453	0.686	0.844	0.897	0.909	0.900	0.889	0.782

The sample period is 1989–2008. Traditional operating accruals are defined as Net Income (Compustat item 172) less Cash from Operations (item 308) divided by average total assets (item 6). Traditional total accruals are computed as Net Income less Net Dividends/Distributions (item 115 + item 127 − item 108) and the change in cash (item 308 + item 311 + item 313) divided by average total assets. Percent Operating Accruals and Percent Total Accruals have the same numerators as Operating Accruals and Total Accruals, respectively, but the denominator for each is the absolute value of Net Income. Return on Assets is Net Income divided by average total assets. Cash Return on Assets is cash from operations divided by average total assets. Total Cash Flow Return on Assets are Net Dividends/Distributions plus the change in cash divided by average total assets. Special items (item 17) are coded as 0 if missing. Mean Three-Year Sales growth is the average annual sales growth in the current year and prior two years. Market Value of Equity is measured at the fiscal year-end (item 25 times item 199). Price (item 199) and Book Value (item 60) are measured at the fiscal year-end. Arbitrage Risk is measured as the standard deviation of the residuals from a regression of firm-specific returns on the CRSP equally weighted index for up to 48 months preceding the month of return portfolio formation. All amounts are winsorized at 1 percent and 99 percent in this table and Table 1 only.

TABLE 3

Overlap between Percent Accruals, Traditional Accruals, and Cash Flow Deciles

Panel A: Percent Operating Accruals

Decile	Traditional Operating Accrual Decile Matches Percent Operating Accrual Decile	Decile Rank of Traditional Operating Accruals	Decile Rank of Cash from Operations Scaled by Total Assets
1	11.78%	3.49	7.07
2	20.38%	3.57	7.05
3	18.06%	3.66	6.76
4	14.75%	3.72	5.98
5	15.14%	3.99	5.54
6	21.14%	4.69	5.43
7	35.67%	5.67	5.24
8	58.33%	7.51	5.11
9	59.66%	9.03	4.40
10	72.95%	9.70	2.44

Panel B: Percent Total Accruals

Decile	Traditional Total Accrual Decile Matches Percent Total Accrual Decile	Decile Rank of Traditional Total Accruals	Decile Rank of Net Dividends and Change in Cash Scaled by Total Assets
1	23.17%	2.36	8.60
2	30.04%	2.01	6.23
3	29.23%	2.29	4.71
4	52.09%	3.57	5.82
5	49.58%	5.54	6.98
6	32.44%	6.73	7.09
7	25.15%	7.46	6.14
8	23.89%	7.91	4.45
9	26.65%	8.35	2.99
10	46.27%	8.80	2.03

Traditional operating accruals are defined as Net Income (Compustat item 172) less Cash from Operations (item 308) divided by average total assets (item 6). Traditional total accruals are computed as Net Income less Net Dividends/Distributions (item 115 + item 127 − item 108) and the change in cash (item 308 + item 311 + item 313) divided by average total assets. Percent Operating Accruals and Percent Total Accruals have the same numerators as Operating Accruals and Total Accruals, respectively, but the denominator for each is the absolute value of Net Income.

To summarize, the lowest deciles of percent accruals are generally composed of different firms than the lowest deciles of traditional accruals—the firms for which the trading strategy will be long. The lowest decile of percent accruals firms are three times larger and much better performing than the lowest decile of traditional accruals, based on either operating or total accruals. The overlap between the percent accruals measures and the traditional accruals measures is much higher in the extreme positive deciles, although still far from perfect.

IV. FUTURE EXCESS RETURNS TO ACCRUAL TRADING STRATEGIES

In this section we document the return to the percent accruals trading strategy and compare it to the traditional accruals strategy and to a cash-from-operations strategy.

Excess Return Computation

Excess returns are computed as buy-and-hold size-adjusted returns, calculated as described in Barber et al. (1999). This method starts by constructing ten size portfolios based on December market value of equity of all NYSE firms. AMEX and NASDAQ firms are then placed in the appropriate size decile using the NYSE breakpoints. Because NYSE firms are typically much larger than AMEX and NASDAQ firms, this results in almost half the observations being sorted into the smallest portfolio. Barber et al. (1999) correct for this skewness by further subdividing the smallest portfolio into four size groups, so that in the end there are 14 size-referent portfolios.[5] For each size portfolio and each month, the returns for each security in the portfolio are compounded over the following 12 months and then averaged across all the securities in the portfolio. The resulting annual return on each reference portfolio constitutes a passive, equally weighted investment in all securities in the portfolio. Excess returns are then the difference between the firm's annual return and the size-matched reference portfolio. For firms that are delisted during the future return period, we calculate the remaining return by taking CRSP's delisting return and then reinvesting the proceeds in the equally weighted reference portfolio. For firms delisted due to poor performance (delisting codes 500 and 520–584), we use a −35 percent delisting return for NYSE/AMEX firms and a −55 percent delisting return for NASDAQ firms, as recommended in Shumway (1997) and Shumway and Warther (1999). Note that the overall mean excess return is zero by construction. For all our results the return window is one year, beginning with the first day of the fourth month after the fiscal year-end.

All significance tests of excess returns are based on Fama-MacBeth t-statistics in order to control for the cross-sectional correlation between firms in a given decile for a given year. In particular, the 19 annual mean excess returns in a decile are averaged and the t-statistic is based on the standard deviation of these 19 observations.

We take extra care to ensure that the returns we report are the result of an implementable trading strategy. For each factor, we use the breakpoints between deciles from the previous year to divide the firms in the current year into ten portfolios. For example, for return windows starting in 1999, we use all observations between August 1998 (four months prior to January 1999) and September 1997 to construct a distribution for the factor and identify the ten breakpoints. This causes our ten portfolios to have slightly different numbers of observations, but makes it absolutely certain that the data necessary to sort the firms into portfolios are available when the return window opens.

Kraft et al. (2006) document a look-ahead bias in many accruals studies. Because many studies were interested in examining the evolution of accruals as well as the stock returns to an accruals-based trading strategy, the sample selection required that the next year's earnings be present. But whether earnings are present in the next year is not known at the time of portfolio formation, so the documented returns are not the result of an implementable trading strategy. Further, the firms with missing future data are not randomly selected; one reason data might be missing in the future is because the firm could have entered bankruptcy. Consistent with this conjecture, Kraft et al. (2006) show that, in a sample of NYSE/AMEX firms, the return to the low-accrual portfolio is 4.2 percent with the bias present but only 1.8 percent without it. In both Kraft et al. (2006) and our study, the excess return in the lowest traditional operating accrual decile is no longer significantly different from zero once the look-ahead bias is removed. The magnitude of the return in the lowest accrual decile is a particularly important part of the evidence supporting

[5] Our results are very similar if we use only ten size portfolios. We prefer the 14-portfolio approach because it provides a closer size match for the smallest firms. Given that many market anomalies are more pronounced for small firms, controlling for size is particularly important.

the accruals anomaly. Without a significantly positive return to the long position in the lowest accrual decile, it is possible that the hedge return no longer exceeds transaction costs, especially given the high transaction costs associated with taking a short position in the high-accrual decile.

Return Results

Table 4 reports the returns by deciles sorted on percent and traditional operating accruals. The hedge return (decile 1 less decile 10) to percent operating accruals is 11.68 percent and is significant at $p < 0.001$. The hedge return comes equally from the long and short position, with an

TABLE 4

Mean Annual Size-Adjusted Returns to Percent Operating Accrual and Traditional Operating Accrual Portfolios

Panel A: Percent Operating Accruals

Decile	Maximum	Size-Adjusted Return	p-value	Number of Obs.
1	−4.04	0.0553	<0.001	8,210
2	−2.04	0.0516	<0.001	8,146
3	−1.34	0.0550	<0.001	8,139
4	−0.96	0.0416	<0.001	8,296
5	−0.67	0.0142	0.416	8,251
6	−0.42	0.0289	0.279	8,182
7	−0.18	0.0112	0.667	8,099
8	0.14	0.0029	>0.500	8,139
9	0.82	−0.0210	0.226	8,109
10		−0.0615	0.009	7,955
Decile 1 − Decile 10		0.1168	<0.001	

Panel B: Traditional Operating Accruals

Decile	Maximum	Size-Adjusted Return	p-value	Number of Obs.
1	−0.2023	0.0127	>0.500	8,237
2	−0.1286	0.0673	0.018	8,100
3	−0.0926	0.0444	0.092	8,197
4	−0.0696	0.0403	0.005	8,234
5	−0.0516	0.0266	0.024	8,040
6	−0.0343	0.0301	0.004	8,313
7	−0.0156	0.0167	0.087	8,155
8	0.0105	−0.0133	0.282	8,156
9	0.0594	−0.0106	0.490	8,127
10		−0.0524	0.029	7,973
Decile 1 − Decile 10		0.0651	0.186	

Returns are the time-series mean annual buy-and-hold size-adjusted returns, calculated as described in Barber et al. (1999) beginning in the fourth month after the fiscal year-end. The p-values are based on two-tailed Fama-MacBeth t-statistics computed over the 19 annual mean returns in a decile and the standard deviation of these 19 observations. The breakpoints between deciles are based on the previous year's cutoffs, so the number of observations in each decile varies slightly. Traditional operating accruals are defined as Net Income (Compustat item 172) less Cash from Operations (item 308) divided by average total assets (item 6). Percent operating accruals have the same numerator but scale by the absolute value of net income.

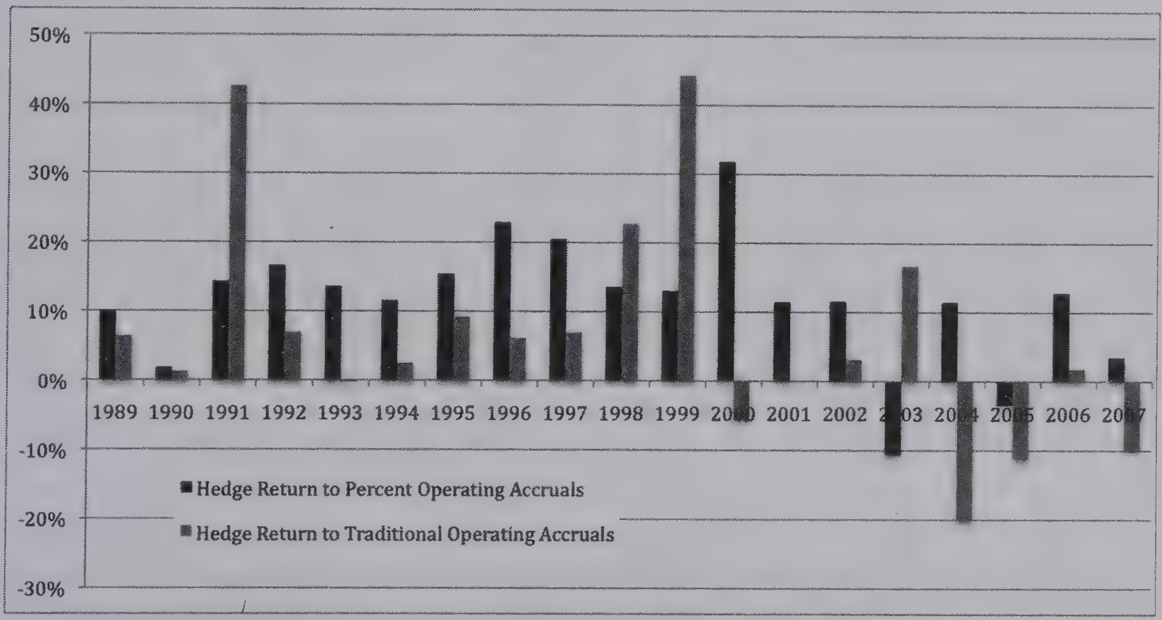
excess return of 5.53 percent in decile 1 and an excess return of −6.15 percent in decile 10; both are significant. In contrast, the hedge return to traditional operating accruals in Panel B is 6.51 percent, but is insignificant with a p-value of 0.186. Further, the excess return in decile 1 is only 1.27 percent and is not significant. Consistent with the fact that the overlap between the two accrual definitions is high for the top decile, the excess returns in decile 10 of traditional operating accruals is −5.24 percent and is significant ($p = 0.029$).[6] Figure 2 compares the hedge returns to percent and traditional operating accrual strategies for each of the 19 years in the sample. The hedge based on percent operating accruals is larger in 15 of the 19 years; a simple binomial test is significant ($p < 0.001$).

Table 5 reports the returns to percent total accruals and traditional total accruals. Panel A shows that the hedge return to percent total accruals is 8.53 percent and significant ($p = 0.001$). The long position in decile 1 contributes 5.49 percent of the hedge return, and is significant ($p = 0.005$). In contrast, Panel B shows that the hedge return to traditional total accruals is not significant. The excess return to the long position is only 1.60 percent and insignificant, although the excess return to decile 10 is −4.23 percent and significant ($p = 0.048$).

In summary, the percent accruals definitions appear to dominate the traditional accrual definitions for both operating and total accruals. Percent accruals generate larger hedge returns, larger excess returns in decile 1 where the two definitions differ the most, and more significant results in all the extreme deciles.

The fact that the accrual anomaly is large and significant when measured as percent accruals offers researchers an empirical regularity that is difficult to reconcile with an efficient stock

FIGURE 2
Annual Size-Adjusted Hedge Returns to Percent and Traditional Operating Accruals (by year in which the 12-month return window begins)



6 There are 967 observations that are in both the first decile of percent operating accruals and the first decile of traditional operating accruals. The mean return for these firms is 7.1 percent, although it is not statistically significant ($p = 0.110$).

TABLE 5

Mean Annual Size-Adjusted Returns to Percent Total Accrual and Traditional Total Accrual Portfolios

Panel A: Percent Total Accruals

Decile	Maximum	Size-Adjusted Return	p-value	Number of Obs.
1	−1.436	0.0549	0.005	7,461
2	−0.890	0.0383	0.110	7,400
3	−0.439	0.0162	>0.500	7,290
4	0.008	0.0387	0.242	7,357
5	0.350	0.0178	0.240	7,193
6	0.658	0.0239	0.035	7,290
7	0.929	0.0076	>0.500	7,225
8	1.291	0.0165	0.332	7,389
9	2.437	−0.0093	0.289	7,264
10		−0.0304	0.092	7,264
Decile 1 − Decile 10		0.0853	0.001	

Panel B: Traditional Total Accruals

Decile	Maximum	Size-Adjusted Return	p-value	Number of Obs.
1	−0.177	0.0160	>0.500	7,418
2	−0.073	0.0604	0.133	7,446
3	−0.027	0.0325	0.028	7,324
4	0.000	0.0420	0.002	7,270
5	0.019	0.0139	0.429	7,227
6	0.037	0.0012	>0.500	7,237
7	0.060	0.0212	0.236	7,272
8	0.095	0.0357	0.005	7,306
9	0.165	−0.0009	>0.500	7,341
10		−0.0423	0.048	7,298
Decile 1 − Decile 10		0.0583	0.313	

Returns are the time-series mean annual buy-and-hold size-adjusted returns, calculated as described in Barber et al. (1999) beginning in the fourth month after the fiscal year-end. The p-values are based on two-tailed Fama-MacBeth t-statistics computed over the 19 annual mean returns in a decile and the standard deviation of these 19 observations. The breakpoints between deciles are based on the previous year's cutoffs, so the number of observations in each decile varies slightly. Traditional total accruals are computed as Net Income less Net Dividends/Distributions (item 115 + item 127 − item 108) and the change in cash (item 308 + item 311 + item 313) divided by average total assets, defined as Net Income (Compustat item 172) less Cash from Operations (item 308) divided by average total assets (item 6). Percent total accruals have the same numerator but scale by the absolute value of net income.

market. It is also inconsistent with the Kothari et al. (2006) agency hypothesis, insofar as the long position contributes a large and statistically significant portion of the excess return. In the next section we explore why extreme values of percent accruals identify mispriced stocks.

V. WHAT DRIVES THE RETURNS TO PERCENT ACCRUALS?

In this section, we explore the source of the excess returns in the portfolios of extreme percent accruals, focusing in particular on the long position in the lowest decile. For brevity, we only consider percent operating accruals, but the results and conclusions are similar for percent total

accruals. Recall that only 12 percent of the observations in the first decile of the traditional operating accruals measure are in the first decile of percent operating accruals. What is the difference between the newly constructed accrual portfolios and the traditional ones, particularly in the lowest accruals portfolio? We offer one explanation and then eliminate a number of other tempting-but-wrong explanations for our return results.

One reason percent operating accruals identifies misvalued stocks is that this measure successfully sorts observations that are more "extreme" into the extreme portfolios than does the traditional operating accruals measure. To understand this, consider the naïve and sophisticated forecasting models posited in the "earnings fixation" hypothesis. Because the scaling variable is what is at issue, we specify each model with unscaled variables and then control for size by including common equity (item 60) as another independent variable, as discussed in Barth and Kallapur (1996), and we estimate the regression on percentile ranks.[7] The naïve model is:

$$\text{Next Year's Net Income} = \alpha_0 + \alpha_1 * \text{Net Income} + \alpha_2 * \text{Common Equity} + \varepsilon.$$

The more sophisticated model distinguishes between cash flows and accruals, resulting in:

$$\begin{aligned}\text{Next Year's Net Income} = \beta_0 + \beta_1 * \text{Cash from Operations} + \beta_2 * \text{Operating Accruals}\\ + \beta_3 * \text{Common Equity} + \varepsilon.\end{aligned}$$

Noting that net income is identically equal to cash from operations plus operating accruals, the expected difference between the predictions of the two models is:

$$\begin{aligned}\text{Difference} = (\beta_0 - \alpha_0) + (\beta_1 - \alpha_1) * \text{Cash from Operations} + (\beta_2 - \alpha_1) * \text{Operating Accruals}\\ + (\beta_3 - \alpha_2) * \text{Common Equity}.\end{aligned}$$

When the difference is positive, the sophisticated model predicts higher next year's net income than the naïve model. Under the earnings fixation hypothesis, this approach identifies a stock as being undervalued—the market behaves as if it is using the naïve model and is subsequently surprised when next year's net income is higher than expected. Consistent with Sloan (1996) and a host of other studies, we show later that $\beta_1 > \alpha_1 > \beta_2$; that is, cash from operations is more persistent than accruals, and the persistence of earnings lies somewhere in between. With this, the most extreme positive differences occur when cash from operations is an extreme positive amount *and* operating accruals is an extreme negative amount. These are exactly the observations that percent operating accruals sorts into the first decile. The large negative accrual results in a large negative numerator and the large positive cash flow of approximately the same magnitude makes the denominator small. In contrast, while the first decile of the traditional accruals contains large negative accruals, it also contains many firms with large negative cash from operations (as seen in Table 2).

To quantify this effect, Table 6, Panel A reports the coefficients from the naïve and sophisticated models, estimating each regression using the percentile ranks of the variables (as in Sloan 1996). To partially control for the look-ahead bias inherent in this type of analysis, if the firm is missing next year's earnings, then it is set equal to the negative of this year's common equity. By writing off all the book value, we are effectively assuming the firm was liquidated. The results in

---

[7] Whether we control for size by including Common Equity in the model has little effect. The coefficient on this variable is almost the same in the naïve and sophisticated models, and so its effect washes out when computing the difference between the models. Consequently, the results in Table 6 are almost identical without this variable.

TABLE 6

Difference between Sophisticated and Naïve Forecasts in Different Accrual Portfolios

Panel A: Percentile Rank Regression of Next Year's Earnings (unscaled)

	Intercept	Current Earnings	Cash from Operations	Operating Accruals	Common Equity
Naïve model	21.96	0.505			0.051
p-value	<0.001	<0.001			<0.001
Sophisticated model	8.86		0.575	0.198	0.048
p-value	<0.001		<0.001	<0.001	<0.001

Panel B: Mean Differences between Sophisticated and Naïve Forecasts by Accrual Decile

Decile	Sorted by Traditional Operating Accruals	Sorted by Percent Operating Accruals
1*	4.6	8.6
2	4.6	5.9
3	2.7	4.3
4	1.3	1.3
5	0.0	−2.4
6	−1.1	−3.6
7	−1.6	−3.1
8	−0.7	−0.8
9	−1.9	−0.8
10*	−8.4	−9.9

* Significantly different using either a two-sample t-test or a Wilcoxon test at the <0.001 level.

For each observation we compute the naive forecast and the sophisticated forecast based on the models in Panel A, and then compute the difference in forecasts. Panel B reports the mean difference in forecasts between the two models for each decile of accruals, sorted on either traditional operating accruals or percent operating accruals. Earnings is data item 172, cash from operations is data item 308, and operating accruals is earnings less cash from operations. Common Equity is data item 60. The dependent variable is next year's earnings. All variables are unscaled. Missing values of the dependent variable are replaced with the negative value of current common equity (item 60).

Table 6 are not sensitive to this assumption—simply eliminating the missing observations yields similar results. As evident in the table, all coefficients are highly significant, and the relative persistence is consistent with prior studies; cash flows are the most persistent, followed by earnings, then followed by accruals.

In Table 6, Panel B we quantify exactly how sorting by percent accruals selects more extreme differences between the naïve and sophisticated models. For each observation, we compute the naïve forecast, the sophisticated forecast, and the difference between the two forecasts. We then report the mean of these differences sorted by percent operating accruals and traditional operating accruals. For instance, in decile 1 of traditional operating accruals, the sophisticated forecast is, on average, 4.6 percent larger than the naïve forecast. By comparison, the mean difference between the forecasts in decile 1 of percent operating accruals is 8.6 percent, almost twice as extreme as the difference for traditional operating accruals. Similarly, the difference in forecasts is −8.4 percent in the tenth decile of traditional operating accruals, as compared to −9.9 percent for the tenth decile of percent operating accruals. The differences in the first decile and tenth decile portfolios are significant at the 0.001 level.

Bradshaw et al. (2001) provide evidence that analysts fail to fully account for the information in accruals in their forecasts. If extreme percent operating accruals are better than traditional accruals at identifying observations where there are extreme differences between the naïve and sophisticated forecasts, then it is possible that they will also do a better job better describing the error in analyst forecasts. To examine this possibility, we regress signed analyst forecast errors of earnings per share on either traditional operating accruals or percent operating accruals. The forecast is the consensus median forecast in the fourth month following the fiscal year-end (the same month that the return window opens) and the realization is for the following fiscal year, both taken from I/B/E/S. Table 7 reports the results. The accruals measures are converted to their decile rank and then scaled to fall in [0, 1]. Consistent with Bradshaw et al. (2001), the coefficient on traditional operating accruals is −0.0039 with a t-statistic of −6.1. Analysts systematically overestimate earnings per share for high-accrual firms and underestimate earnings per share for low-accrual firms. But more importantly, the relation is stronger when percent operating accruals are used, with a coefficient of −0.0049 and a t-statistic of −8.2.

In sum, by conditioning the accrual on the level of net income, percent operating accruals effectively pick out more extreme combinations of cash flows and accruals, resulting in more extreme differences in the naïve and sophisticated forecast models. Consistent with this pattern, analyst forecast errors have a stronger negative association with percent operating accruals than with traditional operating accruals. This evidence supports the earnings fixation hypothesis, insofar as the mispricing of accruals is greatest in places where the hypothesis predicts the greatest differences in naïve and sophisticated beliefs.

Eliminating Some Alternative Explanations for the Success of Percent Accruals

In this section, we run through a number of other candidate reasons for the superior performance of the percent operating accruals.

TABLE 7

Regression of Analyst Forecast Errors on Traditional Operating Accruals and Percent Operating Accruals

	Intercept	Decile Rank of Traditional Operating Accrual	Decile Rank of Percent Operating Accrual
Model 1	−0.0107	−0.0039	
t-statistic	−24.2	−6.1	
p-value	<0.001	<0.001	
Model 2	−0.0102		−0.0049
t-statistic	−25.0		−8.2
p-value	<0.001		<0.001

The forecast error is measured as the actual earnings per share less the median forecast in the fourth month following the fiscal year-end (i.e., the first forecast after the portfolio formation month), taken from I/B/E/S. Traditional operating accruals are defined as Net Income (Compustat item 172) less Cash from Operations (item 308) divided by average total assets (item 6). Percent operating accruals have the same numerator but scale by the absolute value of net income. The independent variables are ranked into deciles, then scaled to be in [0, 1]. The sample size is 40,352.

Percent Accruals Proxies for Cash from Operations

Desai et al. (2004) document a powerful trading strategy based on the ratio of cash from operations to price, and argue that accruals are simply a proxy for this variable, given the strong negative correlation between accruals and cash from operations. They scale by price to capture the well-known value-glamour anomaly (Lakonishok et al. 1994). The value-glamour effect is outside our scope, but in Table 9 of their article, they regress future returns on cash from operations scaled by total assets and accruals scaled by total assets. They find that accruals are no longer significantly related to future returns, but cash from operations remains highly significant. It is therefore possible that percent operating accruals proxy for a more fundamental cash-from-operations anomaly. To examine this possibility, we first document the hedge return to cash from operations scaled by total assets, shown in Table 8. We do not find a significant hedge return, but the return to the long position in the high cash-from-operations portfolio is 6.62 percent and significant at less than 0.001 level. In Table 9, we estimate regressions of next year's returns on the decile rank of percent operating accruals and cash from operations, where the rank is scaled to fall in [0, 1]. In model 1, the coefficient on percent operating accruals is −0.0993, implying that the expected hedge between the extreme deciles is 9.93 percent, and is significant (p = 0.001). In contrast, the coefficient on cash from operations is insignificant in model 2. And finally, when both variables are included in the regression in model 3, the estimated return and significance of percent operating accruals falls slightly, but is still significant, while cash from operations remains insignificant. In sum, sorting on percent operating accruals weakly sorts on cash from operations, as shown in Table 2, but the association with cash flows is not the driving force behind the success of percent operating accruals.

TABLE 8

Mean Annual Size-Adjusted Returns to Cash from Operations Scaled by Total Assets Portfolios

Decile	Maximum	Size-Adjusted Return	p-value	Number of Obs.
1	−0.1522	−0.0496	0.474	8,240
2	−0.0362	0.0030	>0.500	8,043
3	0.0139	0.0026	>0.500	8,041
4	0.0442	0.0145	0.448	8,040
5	0.0681	0.0273	0.049	8,276
6	0.0900	0.0273	0.024	8,092
7	0.1140	0.0378	0.003	8,150
8	0.1435	0.0350	0.009	8,132
9	0.1919	0.0307	0.030	8,251
10		0.0662	<0.001	8,267
Decile 10 − Decile 1		0.1158	0.113	

Returns are the time-series mean annual buy-and-hold size-adjusted returns, calculated as described in Barber et al. (1999) beginning in the fourth month after the fiscal year-end. The p-values are based on two-tailed Fama-MacBeth t-statistics computed over the 19 annual mean returns in a decile and the standard deviation of these 19 observations. The breakpoints between deciles are based on the previous year's cutoffs, so the number of observations in each decile varies slightly. Portfolios are sorted by Cash from Operations (item 308) divided by average total assets (item 6).

TABLE 9

Regressions of Annual Size-Adjusted Returns on Percent Operating Accruals and Cash from Operations

	Intercept	Decile Rank of Percent Operating Accrual	Decile Rank of Cash from Operations Scaled by Total Assets
Model 1	0.0674	−0.0993	
p-value	<0.001	0.001	
Model 2	−0.0212		0.0807
p-value	0.624		0.218
Model 3	0.0346	−0.0784	0.0474
p-value	0.483	<0.001	0.485

Returns are the time-series mean annual buy-and-hold size-adjusted returns, calculated as described in Barber et al. (1999) beginning in the fourth month after the fiscal year-end. The p-values are based on two-tailed Fama-MacBeth t-statistics computed over the 19 annual mean returns in a decile and the standard deviation of these 19 observations. The breakpoints between deciles are based on the previous year's cutoffs, so the number of observations in each decile varies slightly. Percent operating accruals are defined as Net Income (Compustat item 172) less Cash from Operations (item 308), divided by the absolute value of Net Income. The independent variables are ranked into deciles then scaled to be in [0, 1].

Special Items

A common reason a firm has large, negative, unscaled accruals is because it records a non-cash special item, such as an asset write-off, in the period. Dechow and Ge (2006) study this explanation for the traditional accruals measure, reporting that the positive returns in the low-accrual decile in their sample is driven by firms with large negative special items. Indeed, Table 2, Panel B shows that 55.00 percent of the firms in the first decile of traditional accruals have negative special items. In contrast, only 22.83 percent of the firms in the lowest decile of percent accruals have a negative special item. In Table 10, we partition each decile into observations with or without a negative special and examine the returns within each partition.[8]

Table 10, Panel A shows that for both subsamples the hedge return to percent accruals is large and significant; it is 11.88 percent when there is no negative special item and 9.94 percent when a special item is present. Further, the excess return in the first decile is approximately the same size regardless of the special item. Table 10, Panel B shows results consistent with the unpartitioned results for traditional operating accruals; the hedge return is insignificant in both subsamples, as are the returns in the first decile. In sum, the returns to a trading strategy based on percent operating accruals are not particularly sensitive to the presence or absence of a negative special item.

Gains versus Losses

Table 11 bisects the data based on whether the firm reported positive or negative net income. When the data are sorted based on traditional operating accruals, the results are very similar to

[8] We consider two alternative partitions of the sample: (1) the definition in Dechow and Ge (2006) requires that the negative special item exceeds 2 percent of total assets, selecting about 20 percent of the sample, and (2) a definition in the spirit of percent accruals is that the negative special item exceed 25 percent of absolute net income, which also selects about 20 percent of the sample. We prefer using the (unscaled) existence of a negative special item because it yields the same partition regardless of the scale variable used. The conclusions in Table 10 are qualitatively similar using either of these alternative definitions.

TABLE 10

Mean Annual Size-Adjusted Returns for Operating Accruals by Special Item Subsamples

Panel A: Percent Operating Accruals

	No Negative Special Item (64%)			Special Item < 0 (36%)		
Decile	**Return**	**p-value**	**% of sample**	**Return**	**p-value**	**% of sample**
1	0.0511	0.003	5.5%	0.0604	0.002	4.5%
2	0.0414	0.004	5.7%	0.0698	0.002	4.3%
3	0.0449	0.005	5.7%	0.0745	0.001	4.3%
4	0.0446	0.007	5.6%	0.0355	0.076	4.5%
5	0.0112	0.375	5.8%	0.0020	>0.500	4.3%
6	0.0021	>0.500	6.5%	0.0532	0.354	3.5%
7	0.0064	>0.500	6.9%	0.0194	>0.500	3.1%
8	−0.0012	>0.500	7.4%	−0.0116	>0.500	2.6%
9	−0.0272	0.077	7.6%	0.0078	>0.500	2.4%
10	−0.0678	0.019	7.4%	−0.0390	0.171	2.3%
D1 − D10	0.1188	<0.001		0.0994	0.035	

Panel B: Traditional Operating Accrual

	No Negative Special Item (64%)			Special Item < 0 (36%)		
Decile	**Return**	**p-value**	**% of sample**	**Return**	**p-value**	**% of sample**
1	−0.0446	0.125	3.8%	0.0416	0.466	6.3%
2	0.0581	0.006	5.2%	0.0742	0.053	4.7%
3	0.0331	0.153	6.0%	0.0675	0.082	4.1%
4	0.0349	0.000	6.4%	0.0553	0.165	3.7%
5	0.0260	0.023	6.5%	0.0238	>0.500	3.4%
6	0.0244	0.028	6.9%	0.0890	0.042	3.2%
7	0.0158	0.123	7.0%	0.0183	>0.500	3.0%
8	−0.0134	0.301	7.3%	0.0192	>0.500	2.7%
9	−0.0095	>0.500	7.4%	−0.0369	>0.500	2.6%
10	−0.0569	0.028	7.7%	0.0254	>0.500	2.1%
D1 − D10	0.0123	>0.500		0.0162	0.197	

Returns are the time-series mean annual buy-and-hold size-adjusted returns, calculated as described in Barber et al. (1999) beginning in the fourth month after the fiscal year-end. The p-values are based on two-tailed Fama-MacBeth t-statistics computed over the 19 annual mean returns in a decile and the standard deviation of these 19 observations. The breakpoints between deciles are based on the previous year's cutoffs, so the number of observations in each decile varies slightly. Traditional operating accruals are defined as Net Income (Compustat item 172) less Cash from Operations (item 308) divided by average total assets (item 6). Percent operating accruals have the same numerator but scale by the absolute value of net income. Special Items (item 17) are coded as 0 if missing.

those reported by Dopuch et al. (2009), with no significant returns for any decile in the loss subsample, as shown in Panel B. This is a particularly troublesome result for the traditional accrual anomaly, as 34 percent of the observations in the population have losses, and there is nothing in the naïve investor hypothesis that says the result should only hold for firms with positive net income. However, when the data are sorted based on percent operating accruals, Panel A of Table 11 shows that the hedge return and the excess return for the lowest decile are significant in both subsamples, and actually somewhat larger and more significant in the loss subsample than in the gain subsample.

TABLE 11

Mean Annual Size-Adjusted Returns for Operating Accruals by Gain/Loss Subsamples

Panel A: Percent Operating Accruals

	Loss (34%)			Gain (66%)		
Decile	**Return**	**p-value**	**% of Sample**	**Return**	**p-value**	**% of Sample**
1	0.0715	<0.001	3.7%	0.0459	<0.001	6.4%
2	0.0487	0.032	3.7%	0.0533	0.001	6.3%
3	0.0712	0.031	3.8%	0.0408	0.042	6.2%
4	0.0435	0.131	4.4%	0.0379	0.024	5.9%
5	−0.0146	>0.500	4.2%	0.0204	0.224	5.9%
6	0.0143	>0.500	3.8%	0.0233	0.110	6.3%
7	−0.0080	>0.500	3.7%	0.0210	0.075	6.3%
8	−0.0145	>0.500	3.3%	0.0105	>0.500	6.7%
9	−0.0740	0.101	2.1%	−0.0193	0.146	7.9%
10	−0.0636	0.155	1.9%	−0.0634	0.002	7.9%
D1 − D10	0.1351	0.007		0.1093	<0.001	

Panel B: Traditional Operating Accruals

	Loss (34%)			Gain (66%)		
Decile	**Return**	**p-value**	**% of sample**	**Return**	**p-value**	**% of sample**
1	−0.0021	>0.500	8.8%	0.0619	0.128	1.3%
2	0.0539	0.195	5.8%	0.0743	<0.001	4.1%
3	0.0253	>0.500	4.1%	0.0482	0.025	6.0%
4	0.0350	0.361	3.1%	0.0374	0.002	7.0%
5	0.0075	>0.500	2.5%	0.0290	0.137	7.3%
6	0.0506	0.258	2.2%	0.0237	0.132	8.0%
7	0.0357	0.458	2.0%	0.0116	0.411	8.0%
8	−0.0414	0.259	2.0%	−0.0061	>0.500	8.0%
9	−0.0359	0.496	2.0%	−0.0069	>0.500	8.0%
10	−0.0468	0.384	1.9%	−0.0578	0.004	7.8%
D1 − D10	0.0447	>0.500		0.1197	0.009	

Returns are the time-series mean annual buy-and-hold size-adjusted returns, calculated as described in Barber et al. (1999) beginning in the fourth month after the fiscal year-end. The p-values are based on two-tailed Fama-MacBeth t-statistics computed over the 19 annual mean returns in a decile and the standard deviation of these 19 observations. The breakpoints between deciles are based on the previous year's cutoffs, so the number of observations in each decile varies slightly. Traditional operating accruals are defined as Net Income (Compustat item 172) less Cash from Operations (item 308) divided by average total assets (item 6). Percent operating accruals have the same numerator but scale by the absolute value of net income. Loss or Gain is based on the sign of item 172.

Partitioning the percent accruals strategy on gain versus loss offers an interesting connection to Burgstahler and Dichev (1997). In this study, firms with small gains are considered more likely to be manipulating earnings up than firms with small losses. As previously discussed, the extreme deciles of percent accruals are firms with relatively small positive or small negative earnings (making the denominator close to zero). Consistent with the earnings manipulation hypothesis, Table 11, Panel A shows that firms that are in the first decile of percent accruals, and accrue down into a loss position, have higher returns in the subsequent year, as compared to those that accrued down, but into a small gain position. The argument is that their accruals are less suspicious

because they chose to accrue into a small loss. At the other extreme, the earnings manipulation hypothesis suggests that firms in the top decile of percent accruals that accrue up to a small gain are more "manipulative" than those that accrue up to a small loss. The results for this case are not as clear; the returns in the top decile of percent accruals are very similar regardless of the sign of the firm's earnings. However, the significance level of the negative subsequent returns is superior for firms that accrued up to a small positive gain, as compared to those that accrued up to a small loss. Without overstating the significance of these results, there is some evidence that extreme percent accruals combined with the sign of earnings might flag firms that have manipulated earnings in a way that results in predicable market returns in the subsequent year.

Limits to Arbitrage

Mashruwala et al. (2006) find that the accrual anomaly, based on traditional operating accruals, is concentrated in firms with high arbitrage risk, as measured by the idiosyncratic volatility of their stock returns. They also find that it is concentrated in firms with lower share prices, which is a common proxy for transaction costs. They suggest that the combination of risk imposed on arbitrageurs and high transaction costs might leave the accrual anomaly difficult to exploit. To compare the two measures of accruals' exposure to arbitrage risk, we examine the returns to each strategy for each quintile of arbitrage risk. Following Mashruwala et al. (2006), we estimate arbitrage risk as the standard deviation of residuals from a regression of the firm-level returns on the CRSP equally weighted index for up to 48 months preceding the month of return portfolio formation.[9] To keep the portfolio sizes roughly comparable to previous results, we sort percent accruals and traditional accruals into five quintiles. The resulting two-way sort is given in Table 12, and shown graphically in Figure 3.

As seen in Table 12, Panel A, for every quintile of arbitrage risk, the hedge return to percent operating accruals dominates the return to traditional operating accruals. More importantly, the return to percent accruals exceeds 10 percent in the top three quintiles, or over half the sample, and is significant in all but the lowest quintile. Figure 3, Panel B shows the size-adjusted returns to the long position of each strategy across quintiles of arbitrage risk. The returns to the long position for both strategies is essentially zero for the quintile with the lowest arbitrage risk, but for the top three quintiles the percent accruals strategy dominates the traditional accrual strategy, and yields size-adjusted returns of 7 percent, 11 percent, and 8 percent. Table 12, Panel A shows that the long-only returns are significant in the top three quintiles of arbitrage risk, and the hedge returns are significant in the top four quintiles of arbitrage risk. In sum, the results in Figure 3 and Table 12 show that, while percent accruals are not completely insensitive to the level of arbitrage risk, they continue to be large and significant for a much larger portion of the distribution of arbitrage risk than traditional accruals.

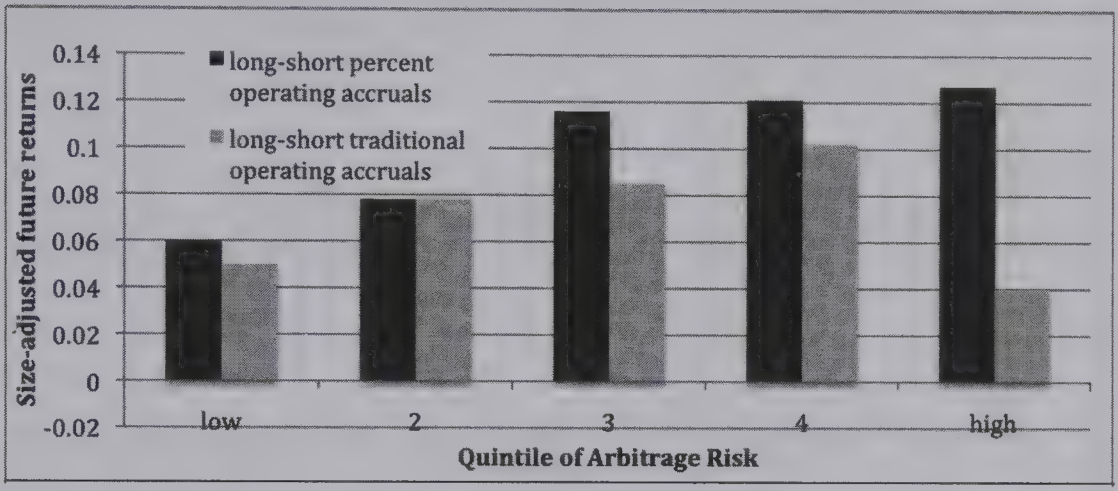
With respect to transaction costs, Panels A and B of Table 2 shows that the firms in the lowest decile of percent operating accruals are considerably larger than the firms in the lowest decile of regular operating accruals. They also have considerably higher share prices. The mean share price in decile 1 for traditional operating accruals is \$7.14 per share, while it is \$13.92 per share for percent operating accruals, almost twice as large. Larger firms and higher share prices have been shown to proxy for lower transaction costs (e.g., Bartov et al. 2000). It is therefore unlikely that greater transaction costs are allowing the higher excess returns to the first decile of percent operating accruals to remain unexploited.

---

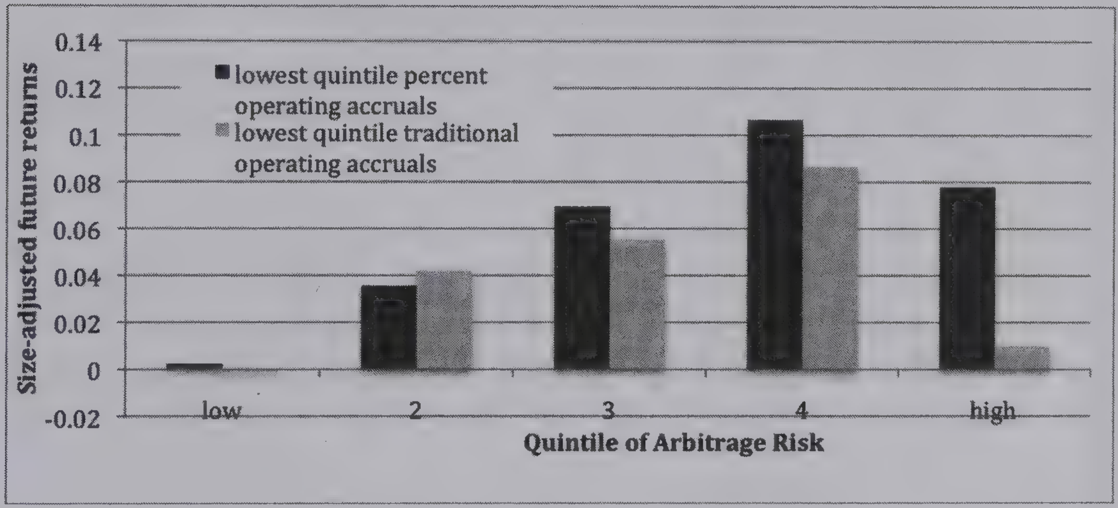
[9] Mashruwala et al. (2006) require that the firm have 48 prior months of returns to be included in the sample. We do not impose this data requirement because it is not generally a requirement imposed on studies of the accrual anomaly. Further, as they report, the returns to an accrual strategy are significantly greater when this data requirement is not imposed. Finally, while estimates based on fewer than 48 observations will be noisier, they will still be unbiased. Requiring 48 months of prior returns, on the other hand, could inadvertently introduce a bias.

FIGURE 3
Annual Size-Adjusted Returns to Percent Operating Accruals by Quintile of Arbitrage Risk

Panel A: Return to Hedge between Top and Bottom Percent Operating Accrual Quintiles



Panel B: Return to Long Position in Lowest Percent Operating Accrual Quintile



Some other candidate explanations for the success of percent operating accruals are as follows:

Greater Association with Growth? The association between accrual measures and "growth" depends on how growth is measured; if it is measured as the increase in operating assets, for instance, then operating accruals and growth are almost synonymous. To assess whether percent accruals are proxying for an underlying "growth" anomaly, without imposing a mechanical relation between the two measures, we compute the average annual sales growth rate over the past three years, ending in the year in which accruals are measured. We use this definition of growth because it is not directly based on changes in operating assets in the same

TABLE 12

Mean Annual Size-Adjusted Returns to Percent and Traditional Accrual Portfolios by Arbitrage Risk Quintiles

Panel A: Percent Operating Accruals

			Quintile of Arbitrage Risk				
			1	**2**	**3**	**4**	**5**
Quintile of Percent Operating Accruals	1	return	0.0024	0.0356	0.0695	0.1056	0.0779
		p-value	>0.500	0.120	0.011	<0.001	0.076
		# of Obs.	3,307	4,114	3,784	3,031	2,126
	2	return	0.0094	0.0308	0.0427	0.0774	0.0817
		p-value	>0.500	0.075	0.012	0.029	0.140
		# of Obs.	4,056	3,371	3,267	3,068	2,673
	3	return	−0.002	0.0259	0.0384	0.053	−0.017
		p-value	>0.500	0.115	0.139	0.250	>0.500
		# of Obs.	3,914	2,881	2,811	2,995	3,832
	4	return	−0.01	−0.003	0.0383	0.0211	−0.005
		p-value	>0.500	>0.500	0.064	>0.500	>0.500
		# of Obs.	3,298	2,967	2,852	3,279	3,842
	5	return	−0.058	−0.042	−0.046	−0.015	−0.049
		p-value	0.041	0.136	0.030	>0.500	0.235
		# of Obs.	1,724	2,979	3,598	3,933	3,830
	Q1 – Q5	return	0.0604	0.0776	0.1155	0.1206	0.1269
		p-value	0.106	0.032	<0.001	0.016	0.034
		# of Obs.	5,031	7,093	7,382	6,964	5,956

(continued on next page)

Panel B: Traditional Operating Accruals

			Quintile of Arbitrage Risk				
			1	2	3	4	5
Quintile of Traditional Operating Accruals	1	return	−0.0010	0.0420	0.0553	0.0865	0.0099
		p-value	>0.500	0.039	0.018	0.065	>0.500
		# of Obs.	1,120	2,157	2,986	4,099	5,975
	2	return	0.0176	0.0335	0.0601	0.0678	0.0310
		p-value	0.459	0.038	0.006	0.130	>0.500
		# of Obs.	3,590	3,823	3,532	3,047	2,439
	3	return	0.0009	0.0283	0.0374	0.0608	0.0535
		p-value	>0.500	0.135	0.063	0.110	0.218
		# of Obs.	5,354	3,675	2,968	2,444	1,912
	4	return	−0.0200	−0.0110	0.0280	0.0265	−0.0100
		p-value	0.461	>0.500	0.074	>0.500	>0.500
		# of Obs.	4,481	3,711	3,251	2,785	2,083
	5	return	−0.0510	−0.0360	−0.0290	−0.0150	−0.0300
		p-value	0.056	0.182	0.194	>0.500	>0.500
		# of Obs.	1,754	2,946	3,575	3,931	3,894
	Q1 – Q5	return	0.0500	0.0780	0.0843	0.1015	0.0399
		p-value	0.213	0.021	0.009	0.076	>0.500
		# of Obs.	2,874	5,103	6,561	8,030	9,869

Returns are the time-series mean annual buy-and-hold size-adjusted returns, calculated as described in Barber et al. (1999) beginning in the fourth month after the fiscal year-end. The p-values are based on two-tailed Fama-MacBeth t-statistics computed over the 19 annual mean returns in a quintile. Arbitrage Risk is measured as the standard deviation of the residuals from a regression of firm-specific returns on the CRSP equally weighted index for up to 48 months preceding the month of return portfolio formation.

period, and because it has been used in prior studies as a measure of "growth" or "glamour" (see Desai et al. 2004, for example). Table 2, Panel A shows how growth varies across percent operating accrual deciles. The table shows only a mild increase in growth across deciles. In the extreme deciles, growth is lower in the first and tenth deciles of percent operating accruals than it is in the first and tenth deciles of traditional operating accruals. As an alternative growth measure, we compute that current year's sales growth. For our sample, the Spearman correlation between the current year's sales growth and either traditional or percent operating accruals is 0.19 and, like the past three-year growth rate, is lower in the extreme deciles of percent accruals than in the extreme deciles of traditional accruals. In sum, while percent operating accruals is weakly correlated with sales growth, it is no more correlated than traditional operating accruals.

Faster Mean Reversion in Accruals? Earlier we reported that the differences between the naïve and sophisticated forecasting models are larger in the extreme deciles of percent operating accruals than in the extreme deciles of traditional operating accruals. This pattern does not lead to a more rapid mean reversion in accruals, however. Forty-one percent of the firms in the first two deciles of traditional operating accruals remain in the first two deciles in the following year. For percent operating accruals, 43 percent remain in the first two deciles in the following year. Thus, while the percent operating accruals measure picks out more extreme differences in forecasts between the naïve and sophisticated models, the accruals for these observations do not reverse any faster than for the traditional operating accruals measure.

Greater Capital Intensity? Another possibility is that, by not scaling by total assets, the percent operating accrual measure is picking up more capital-intensive firms. This is not the case, however. The lowest decile of traditional operating accruals has a ratio of depreciation/amortization to average total assets of 0.1261, while the same measure is only 0.0684 for the lowest decile of percent operating accruals.

More Extreme "Outlier" Returns? Kraft et al. (2006) show that a large part of the positive excess return in the low-accrual portfolio is due to a few firms with extreme positive returns. The 99th percentile of excess returns in the first decile of traditional operating accruals is 5.23, but is only 2.87 in the first decile of percent operating accruals. Further, the mean of the top five excess returns is almost twice as high in the first decile of traditional operating accruals than it is in the first decile of percent operating accruals. While the distribution of returns is clearly skewed in both samples, the percent operating accrual measure is less sensitive to extreme returns than is the traditional measure. Incidentally, of the 83 most extreme excess returns in the top percentile of the first decile of traditional operating accruals and percent operating accruals, only 4 observations are common between the two portfolios.

VI. CONCLUSION

Accruals are the main work product of financial accounting. It is therefore vitally important that we understand how accruals relate to equity valuation. This exploration goes beyond attempts to refine an investment strategy and produce ever-bigger excess returns. By documenting another means of using accrual information to identify misvalued stocks, we learn more about what the market does and does not understand.

We offer a subtle innovation. By redefining accruals to be relative to income rather than to total assets—that is, percent accruals—we offer a new measure of accruals that selects radically different firms and produces excess returns in subsamples of the population for which the traditional accruals measure performs poorly. Our results are consistent with the earnings fixation hypothesis insofar as extreme values of percent accruals correspond with cases in which the

difference between a naïve and sophisticated forecasting model is the most pronounced. As the literature continues to investigate the underlying explanation for the accruals anomaly, our results provide additional regularities that the theory must accommodate.

REFERENCES

Barber, B., J. Lyon, and C. Tsai. 1999. Improved methods for tests of long-run abnormal stock returns. *The Journal of Finance* 54 (February): 165–201.

Barth, M. E., and S. Kallapur. 1996. The effects of cross-sectional scale differences on regression results in empirical accounting research. *Contemporary Accounting Research* 13 (2): 527–567.

Bartov, E., S. Radhakrishnan, and I. Krinsky. 2000. Investor sophistication and patterns in stock returns after earnings announcements. *The Accounting Review* 75 (1): 43–63.

Beaver, W., M. McNichols, and R. Price. 2007. Delisting returns and their effect on accounting-based market anomalies. *Journal of Accounting and Economics* 43 (2–3): 341–368.

Beneish, M., and M. Vargus. 2002. Insider trading, earnings quality and accrual mispricing. *The Accounting Review* 77 (4): 755–791.

Bradshaw, M., S. Richardson, and R. Sloan. 2001. Do analysts and auditors use information in accruals? *Journal of Accounting Research* 39 (1): 45–74.

Burgstahler, D., and I. Dichev. 1997. Earnings management to avoid earnings decreases and losses. *Journal of Accounting and Economics* 24 (1): 99–126.

Cheng, A., and W. Thomas. 2005. The effects of alternative models and estimation procedures on the abnormal accrual anomaly. Working paper, University of Houston.

Collins, D., and P. Hribar. 2002. Errors in estimating accruals: Implications for empirical research. *Journal of Accounting Research* 40 (March): 105–134.

D'Avolio, G., E. Gildor, and A. Shleifer. 2002. *Technology, Information Production, and Market Efficiency: Economic Policy for the Information Economy*. A Symposium Sponsored by the Federal Reserve Bank of Kansas City, KS.

Dechow, P., and W. Ge. 2006. The persistence of earnings and cash flows and the role of special items: Implications for the accrual anomaly. *Review of Accounting Studies* 11 (2-3): 253–296.

Dopuch, N., C. Seethamraju, and W. Xu. 2009. The pricing of accruals for profit and loss firms. *Review of Quantitative Finance and Accounting*. 34, (4): 505–516.

Desai, H., S. Rajgopal and M. Venkatachalam. 2004. Value-glamour and accruals mispricing: One anomaly or two? *The Accounting Review* 79 (2): 355–85.

Fairfield, P., S. Whisenant, and T. Yohn. 2003. Accrued earnings and growth: Implications for future profitability and market mispricing. *The Accounting Review* 78 (1): 353–371.

Houge, T., and T. Loughran. 2000. Cash flow is king? Cognitive errors by investors. *The Journal of Psychology and Financial Markets* 1 (March): 161–175.

Kothari, S., E. Loutskina, and V. Nikolaev. 2006. Agency theory of overvalued equity as an explanation for the accrual anomaly. Working paper, Massachusetts Institute of Technology.

Kraft, A., A. Leone, and C. Wasley. 2006. An analysis of the theories and explanations offered for the mispricing of accruals and accrual components. *Journal of Accounting Research* 44 (2): 297–339.

Lakonishok, J., A. Shleifer, and R. Vishny. 1994. Contrarian investment, extrapolation, and risk. *The Journal of Finance* 49 (5): 1541–1578.

Lesmond, D., and W. Wang. 2006. The long and short of accrual based trading strategies. Working paper, Tulane University.

Mashruwala, C., S. Rajgopal, and T. Shevlin. 2006. Why is the accrual anomaly not arbitraged away? The role of idiosyncratic risk and transaction costs. *Journal of Accounting and Economics* 42 (1-2): 3–33.

Richardson, S., R. Sloan, M. Soliman, and I. Tuna. 2005. Accrual reliability, earnings persistence and stock prices. *Journal of Accounting and Economics* 39 (3): 437–485.

Shumway, T. 1997. The delisting bias in CRSP data. *The Journal of Finance* 52 (1): 327–340.

———, and V. A. Warther. 1999. The delisting bias in CRSP's Nasdaq data and its implications for the size effect. *The Journal of Finance* 54 (6): 2361–2379.

Sloan, R. 1996. Do stock prices fully reflect information in accruals and cash flows about future earnings? *The Accounting Review* 71 (July): 289–316.

Zach, T. 2004. Evaluating the "accrual-fixation" hypothesis as an explanation for the accrual anomaly. Working paper, Washington University.

THE ACCOUNTING REVIEW
Vol. 86, No. 1
2011
pp. 237–258

American Accounting Association
DOI: 10.2308/accr.00000014

Do Investors Understand *Really* Dirty Surplus?

Wayne R. Landsman
The University of North Carolina at Chapel Hill

Bruce L. Miller
University of California, Los Angeles

Ken Peasnell
Lancaster University

Shu Yeh
National Taiwan University

ABSTRACT: This study addresses whether firms' share prices correctly reflect two accounting measures: dirty surplus and really dirty surplus. Dirty surplus is readily observable from the financial statements, but really dirty surplus, which arises from recognizing equity transactions such as employee stock option exercises at other than fair market value, is not. Findings show that dirty surplus and really dirty surplus are irrelevant for forecasting abnormal comprehensive income. However, findings also indicate that investors appear to undervalue really dirty surplus. Hedge returns are insignificant when portfolios are formed based on dirty surplus, but are significantly positive based on really dirty surplus. Really dirty surplus positive hedge returns are robust to a variety of sensitivity tests. Taken together, the findings are consistent with either investors over-valuing firms that have large negative really dirty surplus or really dirty surplus being correlated with an unmodeled risk factor.

Keywords: *dirty surplus accounting; equity valuation; hedge returns.*

Data Availability: *Data are available from sources identified in the paper.*

We thank workshop participants at the 2009 European Accounting Association Congress, 2008 UNC-Duke Fall Camp, 2008 Yale Fall Camp, Chinese University of Hong Kong, National Taiwan University, National University of Singapore, Norwegian School of Economics, University of California, Los Angeles, University of Graz, and Vienna University of Economics and Business Administration, as well as Frøystein Gjesdal and two anonymous reviewers for helpful comments. We acknowledge funding from the Center for Finance and Accounting Research at The University of North Carolina at Chapel Hill and National Science Council, Republic of China (NSC 96-2416-H-002-038; NSC97-2410-H-002-042).

Editor's note: Accepted by Sanjay Kallapur.

Submitted: July 2008
Accepted: July 2010
Published Online: January 2011

I. INTRODUCTION

A substantial and growing literature considers whether investors properly assess the characteristics of earnings and its components when setting stock prices. The question we address is whether firms' share prices correctly reflect two accounting measures that have received relatively little attention to date. The first of these is commonly referred to as "dirty surplus," which is a component of comprehensive income that is excluded from reported earnings, and therefore violates clean surplus accounting. We label the second accounting measure we consider "really dirty surplus," which arises when a firm issues or reacquires its own shares in a transaction that does not record the shares at fair market value. Examples of this kind of transaction are shares issued in a stock option exercise and a conversion of a bond into common stock. Prior to the implementation of FASB Statement No. 141, the pooling-of-interests method of accounting for business combinations could also result in substantial really dirty surplus. If investors fully understand the predictive value of these accounting amounts, then it should not be possible to develop a profitable trading strategy based on the magnitudes of these items.

Unlike dirty surplus, which is readily observable from the financial statements, really dirty surplus is unobservable. That is, even the most sophisticated investor cannot estimate readily the valuation impact of equity transactions that give rise to really dirty surplus because equity transactions are recognized in the financial statements using an accounting-based rather than a market-based measure of the value of equity. The estimation task investors face is exacerbated by the fact that a firm can engage in numerous such transactions throughout the year. As a result, really dirty surplus transactions are less likely to be correctly priced than are dirty surplus transactions.

Using financial statement and stock price data from 1976–2006, we first assess whether investors properly value each of these two components of earnings by estimating a residual income forecasting equation and an attendant valuation equation that includes both of these components. If current residual income is sufficient for forecasting the next period's residual income and current residual income and equity book value are sufficient for valuing current equity, then the forecasting and valuation coefficients on the income components of interest will be linked in a predictable manner. Finding a mismatch between the components' forecasting and valuation equation coefficients would be consistent with investors' mispricing of the components.

We also conduct hedge portfolio returns tests. We adopt a buy-and-hold strategy to go long in firms with relatively large dirty or really dirty surplus and to go short in firms with relatively small dirty or really dirty surplus. We conjecture that small firms' prices are less likely to reflect fully all publicly available information because investors incur proportionately greater transaction costs; as a result, they are less closely followed and less likely to be subject to detailed accounting analysis. We therefore conduct both sets of tests separately for small, medium, and large firms to assess whether pricing effects are related to firm size.

We find that both dirty surplus and really dirty surplus are irrelevant for forecasting abnormal very comprehensive income for all three firm-size groups. Taking these results at face value, if investors correctly understand the implications of these persistence findings for valuation, then each kind of dirty surplus should be irrelevant for valuation for all firms. This prediction is borne out in the case of dirty surplus. However, the findings indicate that investors appear to undervalue really dirty surplus, which is consistent with investors being unable to assess the economic implications of really dirty surplus transactions.

Buy-and-hold hedge return results support the findings from the tests linking the forecasting and valuation equations. As expected, hedge returns are insignificantly different from zero when based upon dirty surplus, regardless of firm size and investing horizon. In contrast, the hedge returns based on really dirty surplus are significantly positive for all three firm-size groups. We also consider an alternative to our buy-and-hold procedure for computing hedge returns. Findings based on mean returns for monthly calendar-time hedge portfolios indicate that significantly posi-

tive hedge returns are concentrated within small firms. Findings from additional tests reveal that inferences relating to hedge returns are insensitive to including controls for four previously identified mispricing anomalies, and to sampling procedures designed to attempt to focus on sources of really dirty surplus.

Taken together, the findings are consistent with investors over-valuing firms that have large negative really dirty surplus. However, several cautionary notes are in order. First, although the hedge returns findings are consistent with mispricing of really dirty surplus, the possibility remains that the mismatch of the really dirty surplus forecasting and valuation coefficients is the result of model misspecification rather than mispricing. Second, as is likely the case with investors, we are unable to trace the sources of really dirty surplus to particular types of equity transactions. As a result, we cannot determine the extent to which potential mispricing arises from each type of transaction, i.e., our findings can only be interpreted as reflecting the aggregate effect of the various types of transactions. However, even if we could trace the sources of really dirty surplus, any resulting hedge returns might still be attributable to an unmodeled risk factor.

Our study adds to prior research finding evidence of investors' apparent failure to link the forecasting attributes of accounting amounts with the pricing implications (e.g., Bernard and Thomas 1989; Sloan 1996; Barth et al. 1999; Bradshaw and Sloan 2002; Burgstahler et al. 2002; Brown and Sivakumar 2003; Doyle et al. 2003; Landsman et al. 2007). Our findings support prior studies that find that investors understand the forecasting properties and valuation implications of dirty surplus (Dhaliwal et al. 1999; O'Hanlon and Pope 1999; Biddle and Choi 2006; Chambers et al. 2007). Although Landsman et al. (2006) examine valuation implications of expected future equity transactions arising from the exercise of employee stock options, their study does not address whether investors take full account of the valuation implications of past option exercises. Core et al. (2002) report findings suggesting that dilutive transactions, including those arising from employee stock option grants, are poorly dealt with in reported diluted earnings per share, leaving open the possibility that investors may have difficulty in valuing such transactions.

Section II provides the motivation for the study and explains how dirty surplus and really dirty surplus are defined. Section III presents the research design, including computation of dirty surplus and really dirty surplus, development of the forecasting and valuation equations, and description of our hedge return strategy. Section IV describes the sample and data, and Section V presents the findings. Section VI summarizes and concludes.

II. MOTIVATION

The empirical issue that is central to this research is whether firms' share prices correctly reflect two accounting measures that have received relatively little attention to date. The first of these is commonly referred to as *dirty surplus* (*DS*), which is a component of comprehensive income that is excluded from reported earnings and therefore violates clean surplus accounting (Ohlson 1995; Feltham and Ohlson 1995). Dirty surplus accounting results in the basic residual income valuation model yielding an inaccurate estimate of equity value because the sum of current book value and future net incomes does not equal the sum of future net dividends.

The second accounting measure we consider we label *really dirty surplus* (*RDS*), which arises when a firm issues or reacquires its own shares in a transaction that does not record the shares at fair market value. The primary sources of *RDS* include employee stock option exercises, conversion of preferred stock and bonds, and mergers accounted for as pooling-of-interests. Whereas equity issued under employee stock option exercises or convertible instruments can give rise to unrecorded expenses, equity issued under pooling-of-interests gives rise to an unrecorded asset.[1]

[1] Equity issued under the pooling-of-interests method is not recognized at fair market value. In contrast, if purchase

RDS violates the *super-clean surplus* concept (Feltham 1996; Christensen and Feltham 2003), under which it is assumed that share issuances are recorded at fair market value. When this condition is violated, the discounted present value of future net dividends (or equivalently, the sum of equity book value and discounted present value of future abnormal earnings) will not equal the market value of equity relating to current shares outstanding, but rather will equal the market value of equity relating to current shares outstanding plus the market value of other equity claimants. Because the equity transactions that give rise to *RDS* generally are recorded in the financial statements at less than market value, *RDS* is generally negative.[2]

DS is readily observable from the financial statements. When one takes a *clean surplus accounting* perspective, as comprehensive income does, *dirty surplus* becomes a component of earnings. Dirty surplus is conventionally defined as the sum of recognized revenue or expense items that bypass the income statement. Unlike *DS*, *RDS* is not reported in the financial statements. A *super-clean surplus accounting* perspective requires that both *dirty surplus* and *really dirty surplus* become components of earnings so that the discounted stream of future residual incomes and current equity book sums to the market value of equity of current shareholders. If investors fully understand the implications of *DS* and *RDS* for valuation, then it should not be possible to develop a profitable trading strategy based on the magnitudes of these items.

To see these points more clearly, consider first the following version of clean surplus accounting:

$$BVE_t = BVE_{t-1} + X_t + DS_t - Div_t + P_t^A(N_t - N_{t-1}), \tag{1}$$

where BVE_t is defined as ending equity book value, X_t represents net income, DS_t is dirty surplus, Div_t is dividends, N_t is the number of shares outstanding at the end of period t, and P_t^A is the price per share used to record the issuance or reacquisition of equity shares in the accounting system. Note that if DS_t is zero, then the accounting is said to satisfy clean surplus accounting.

Let P_t^M be the market price per share at the date of issuance or reacquisition of equity shares. We define really dirty surplus, RDS_t, by:

$$RDS_t \equiv (N_t - N_{t-1})(P_t^A - P_t^M). \tag{2}$$

By combining Equation (1) and Equation (2), we arrive at:

$$DS_t + RDS_t = BVE_t - BVE_{t-1} - X_t + Div_t - P_t^M(N_t - N_{t-1}). \tag{3}$$

DS_t, BVE_t, BVE_{t-1}, X_t, and Div_t are readily observable in the financial statements. The final term on the right-hand side of Equation (3), $P_t^M(N_t - N_{t-1})$, is not reported in the financial statements and therefore needs to be estimated.

Note that if both DS_t and RDS_t are zero in Equation (3), then the accounting is said to satisfy super-clean surplus accounting. The next section allows for both nonzero DS_t and RDS_t and super-clean surplus accounting by setting $VCNI_t = X_t + DS_t + RDS_t$, where *VCNI* is "very comprehensive" net income. Our definition of *RDS*, and hence *VCNI*, attributes all of the violation of super-clean surplus accounting to the period during which the equity transaction is recorded at a price other than fair market value. Christensen and Feltham (2003) show that when super-clean surplus accounting holds in periods subsequent to time t, application of the residual income

accounting was applied instead of the pooling-of-interests method, *RDS* would not arise; instead, the amount of *RDS* attributed to pooling-of-interests would be recognized in the financial statements as goodwill.

2 It is possible that *RDS* could be positive. For example, consider a bond with a book value of \$175 and fair value of \$125 that is converted into equity whose fair value is \$150. In this case, *RDS* would be a positive amount equal to \$25. Likewise, unrecorded goodwill associated with a merger accounted for under pooling-of-interests could give rise to negative unrecorded goodwill and, hence, positive *RDS*.

valuation model will yield an estimate of equity value that equals the market value of equity of existing shares. Whether super-clean surplus accounting holds up to and including period t simply affects the opening balance of equity book value at time t. However, if P^M and P^A differ for transactions in periods subsequent to time t, then super-clean surplus accounting will be violated and, hence, the residual income valuation model will not yield an estimate of equity value that equals the market value of existing shares.[3]

As stated above, the empirical issue that is central to this research is whether firms' price per share correctly reflects *DS* and *RDS*. If it does, then one should not be able to develop a trading strategy based on *DS* or *RDS* that generates future abnormal returns. There are several reasons why we expect that *RDS* is the better earnings component on which to base a trading strategy. First, as noted above, unlike *DS*, *RDS* is not reported in the financial statements. Second, *RDS* appears to be inherently complex. For example, for most earnings components, any "overstatement" or "understatement" reverses in future periods; this does not hold for *RDS*. Third, research on *DS* (Dhaliwal et al. 1999; O'Hanlon and Pope 1999; Biddle and Choi 2006; Chambers et al. 2007) is not especially encouraging about the possibility that it can be used to construct a profitable trading strategy.

Nonetheless, there are at least two compelling reasons for conducting our tests for *DS* as well as *RDS*. First, our study is the first to examine whether investors properly price *DS* based on the forecasting and valuation equations in the Ohlson (1999) model as well as on hedge return tests. Second, because we do not necessarily expect to find evidence of mispricing relating to *DS*, finding this is the case mitigates concerns that finding evidence of mispricing relating to *RDS* is attributable to misspecification of our empirical procedures.

III. RESEARCH DESIGN

Computation of *DS* and *RDS*

Following Dhaliwal et al. (1999) and Chambers et al. (2007), we compute *DS* as the sum of (1) the change in the balance of unrealized gains or losses on marketable securities (change in Compustat #238), (2) the change in the cumulative foreign exchange adjustment (change in Compustat #230), and (3) 0.65 times the change in additional pension liability in excess of unrecognized prior service costs (change in Min [(Compustat #297 − #298), 0]).

Based on Equation (3), we compute *RDS* as the change in the book value of common equity (Compustat #60 + #227 − #242), less *DS*, less net income (Compustat #172 − #19), plus dividends (Compustat #21), less share price at middle of fiscal year times change in common shares outstanding (Compustat #25, adjusted for stock dividends and splits).[4] Note that because we (and investors) cannot readily compute using the individual underlying equity transactions, *RDS* likely measures the true underlying construct with error.[5] Share prices are from the CRSP database.

---

[3] Landsman et al. (2006) show that, in the case of employee stock options (i.e., contingent equity), when only clean surplus holds, the estimate of equity value equals the sum of the market value of existing shares and employee stock options. The study's model considers the case in which employee stock options are granted at time t or earlier. The residual income valuation model does not yield an estimate of the value of existing shares because the options are not yet exercised, and when they are exercised in the future, the new shares will be recognized at P^A rather than P^M.

[4] To the extent that our definition of *DS* does not include all dirty surplus items (e.g., the cumulative effects on equity of retrospective accounting changes), *DS* will be measured with error. Because *RDS* is net of *DS*, such items will appear as part of our measure of *RDS*. Also, our measure of *RDS* includes treasury stock transactions taking place at prices that differ from the market price at the middle of the fiscal year.

[5] In particular, the use of mid-year prices in the construction of *RDS* is arbitrary. We test the sensitivity of our findings to measuring *RDS* at alternative dates using both end-of-year and average of beginning- and end-of-year prices. Untabulated findings based on these alternative measures reveal that none of the inferences are affected.

Forecasting and Valuation Equations

To examine how the dirty surplus and really dirty surplus components of income relate to equity value, we adopt the abnormal earnings forecasting and equity valuation equations from Barth et al. (1999), which are based on the linear information system developed in Ohlson (1999):

$$VCNI^a_{it+1} = \omega_0 + \omega_1 VCNI^a_{it} + \omega_2 DS_{it} + \omega_3 RDS_{it} + \omega_4 BVE_{it} + \varepsilon_{it+1}. \quad (4)$$

$$MVE_{it} = \alpha_0 + \alpha_1 VCNI^a_{it} + \alpha_2 DS_{it} + \alpha_3 RDS_{it} + \alpha_4 BVE_{it} + u_{it}. \quad (5)$$

Equation (4) is the abnormal earnings forecasting equation, where abnormal very comprehensive earnings, $VCNI^a_t$, is defined as very comprehensive earnings, $VCNI_t$, less a normal return on beginning equity book value, BVE_{t-1}, i.e., $VCNI_t - rBVE_{t-1}$. Very comprehensive income is net income, NI_t, plus both dirty surplus and really dirty surplus. Following Ohlson (1999) and Barth et al. (1999), $VCNI_t$ is partitioned into NI_t, DS_t, and RDS_t. The linear information system represented by Equation (4) and Equation (5) implicitly assumes that current earnings amounts are predictive of all future earnings. To the extent that this assumption is violated, the algebraic links between forecasting coefficients in Equation (4) and the valuation coefficients in Equation (5) described below do not necessarily hold. Of particular significance to this study is whether current realizations of *DS* and *RDS* are predictive of future $VCNI^a$ that includes future realizations of these variables.

In Equation (4), ω_1 reflects the persistence of abnormal earnings. Prior research (e.g., Dechow et al. 1999; Barth et al. 1999, 2005) leads us to predict that ω_1 is positive.[6] The coefficients on the *DS* and *RDS* earnings components, ω_2 and ω_3, reflect the incremental effects on the forecast of abnormal earnings of knowing these components. If all earnings components have the same ability to forecast $VCNI^a_{t+1}$, then ω_2 and ω_3 will both equal zero; thus, knowing each component of earnings does not aid in forecasting abnormal earnings. As a result, we test the null hypotheses that $\omega_2 = 0$ and $\omega_3 = 0$ against the alternative that $\omega_2 \neq 0$ and $\omega_3 \neq 0$.

Following Ohlson (1999, 150), we define *DS* (*RDS*) as being "forecasting-irrelevant" if the quadruple $\{NI_t, RDS_t, BVE_t, BVE_{t-1}\}$ ($\{NI_t, DS_t, BVE_t, BVE_{t-1}\}$) contains the same information as the quintuple $\{NI_t, DS_t, RDS_t, BVE_t, BVE_{t-1}\}$ for purposes of forecasting $VCNI^a_{t+1}$. Because *DS* and *RDS* are components of $VCNI^a_t$, the total coefficients on DS_t and RDS_t are $\omega_1 + \omega_2$ and $\omega_1 + \omega_3$. ω_4 is not included in the total coefficient on either DS_t or RDS_t because BVE_t is unchanged across the different definitions of clean surplus and is therefore invariant to the definition of clean surplus. Thus, if $\omega_1 + \omega_2 = 0$ ($\omega_1 + \omega_3 = 0$), *DS* (*RDS*) is irrelevant for forecasting abnormal earnings. Conversely, if $\omega_1 + \omega_2 \neq 0$ ($\omega_1 + \omega_3 \neq 0$), then *DS* (*RDS*) is said to have abnormal earnings "forecasting relevance." To examine whether dirty surplus and really dirty surplus components of comprehensive income are forecasting-irrelevant, we test the null hypotheses that $\omega_1 + \omega_2 = 0$ and $\omega_1 + \omega_3 = 0$ against the alternatives that $\omega_1 + \omega_2 \neq 0$ and $\omega_1 + \omega_3 \neq 0$. Note that ω_1 reflects the forecasting relevance of the $VCNI^a_t - DS_t - RDS_t = NI_t - rBVE_{t-1}$ component of $VCNI^a_t$.

Equation (5) is the valuation equation based on the information dynamics in Equation (4). α_2 and α_3, the valuation multiples on *DS* and *RDS*, can be interpreted in a symmetrical fashion. This follows from the fact that although DS_t is by definition included in BVE_t, it follows from Equation (1) and Equation (3) that RDS_t that arises from dilutive transactions is normally included in BVE_t

[6] Ohlson (1995, 1999) permits the forecasting and valuation equations to include "other information." Fairfield et al. (2003) show that accruals and asset growth have incremental ability to predict future return on assets. Accordingly, viewing accruals and asset growth as "other information," below we report findings from alternative specifications of Equation (4) and Equation (5) that include proxies for these variables as additional explanatory variables.

as well.[7] Analogous to the interpretation of ω_2 (ω_3) in Equation (4), a_2 (α_3) reflects the incremental effect on valuation from knowing *DS* (*RDS*). If all earnings components are equally persistent, then they should have the same relation with equity value. If this is the case, then a_2 and a_3 will equal zero, and knowing each component of earnings will not aid in explaining equity value. Thus, we test the null hypothesis that $a_2 = 0$ ($a_3 = 0$) against the alternative that $a_2 \neq 0$ ($a_3 \neq 0$). We define *DS* (*RDS*) as being "valuation-irrelevant" if the quadruple $\{NI_t, RDS_t, BVE_t, BVE_{t-1}\}$ ($\{NI_t, DS_t, BVE_t, BVE_{t-1}\}$) contains the same information as the quintuple $\{NI_t, DS_t, RDS_t, BVE_t, BVE_{t-1}\}$ for purposes of valuation. Also analogous to Equation (4), the total valuation coefficient on *DS* (*RDS*) equals $a_1 + a_2$ ($a_1 + a_3$). Thus, if $\alpha_1 + \alpha_2 = 0$ ($\alpha_1 + \alpha_3 = 0$), *DS* (*RDS*) is irrelevant for valuation.[8] Conversely, if $a_1 + a_2 \neq 0$ ($a_1 + a_3 \neq 0$), then *DS* (*RDS*) is "valuation-relevant." Analogous to the interpretation of ω_1 in Equation (4), a_1 reflects the value relevance of the $VCNI_t^a - DS_t - RDS_t = NI_t - rBVE_{t-1}$ component of $VCNI_t^a$.

Barth et al. (1999) derive a formula linking the coefficients in Equation (4) and the two suppressed equations with the coefficients in Equation (5). For our purposes, we are not interested in exact coefficient magnitudes based on imposing a full set of linear information dynamics. Instead, we are interested in the weaker prediction that the sign of $a_1 + a_2$ ($a_1 + a_3$) will be based on the sign of $\omega_1 + \omega_2$ ($\omega_1 + \omega_3$).

If prices are determined rationally, then if *DS* or *RDS* is irrelevant for forecasting the next period amount, each should be valuation-irrelevant as well if the linear dynamics in Equation (4) and Equation (5) hold. Also, the sign of $a_1 + a_2$ ($a_1 + a_3$) will be the same as the sign of $\omega_1 + \omega_2$ ($\omega_1 + \omega_3$). If we find apparent evidence of mispricing based on the empirical coefficients from estimating Equation (4) and Equation (5), then a buy-and-hold strategy of going long in relatively underpriced stocks and short in relatively overpriced stocks should yield excess returns.

Any mismatch between the forecasting and valuation results for *DS* or *RDS* need not necessarily be attributable to its mispricing by investors. It might be the case, for example, that although RDS_t cannot be used to forecast $VCNI_{t+1}^a$, it could be used to forecast $VCNI_{t+k}^a$ ($k = 2,3,\ldots$). If this were the case, then *RDS* would not be valuation-irrelevant, and the mismatch between forecasting and valuation coefficients would be attributable to variables omitted from both equations. The hedge returns tests provide a means of examining this issue.

If transaction cost considerations imply that small firms are more difficult to price, then we would expect the hedge portfolio returns to be greater than in the case of larger firms. Therefore, we estimate and test predictions relating to Equation (4) and Equation (5) separately for small, medium, and large firms based on equity market value and conduct hedge return tests also separately for small, medium, and large firms.

---

[7] This can be illustrated by the following simple bond conversion example. Consider a firm that has a convertible bond outstanding on its books at \$100 that is converted into shares worth \$150 at time *t*. Under current GAAP, the share issuance will be recorded at \$100. If we assume for simplicity that $X_t = DS_t = 0$ and that $BVE_{t-1} = \$1{,}000$, it follows that $BVE_t = 1{,}000 + 100 = \$1{,}100$. If this transaction were to be accounted for on a super-clean surplus basis, the share issuance would be recorded at \$150, with the resultant cost of conversion appearing as $RDS_t = -50$. We can deduce from Equation (3) that under super-clean surplus accounting $BVE_t = 1{,}000 - 50 + 150 = \$1{,}100$ as well. Although the calculations are more complex in the case of employee stock options, the same conclusion applies. Note that in the case of mergers accounted for under pooling-of-interests, the inclusion of RDS_t in BVE_t would leave BVE_t unchanged only if the asset (goodwill) associated with an acquisition was immediately expensed.

[8] Note that under the Ohlson (1999) framework, value irrelevance (e.g., which occurs for *DS* (*RDS*) when $\alpha_1 + \alpha_2 = 0$ ($\alpha_1 + \alpha_3 = 0$)) of an earnings component implies that it has no impact on goodwill, which is the difference between equity market value and book value. Ohlson (1999, 152) further states that "an incremental dollar of transitory earnings adds a dollar to market value. This claim is easy to validate as long as one keeps in mind that a dollar of transitory earnings also adds a dollar to book value."

Hedge Portfolio Strategy and Procedure

Hedge Strategy Overview

We determine the hedge portfolio strategy in the following manner. First, for each sample year, we rank firms according to either *DS* or *RDS* as a fraction of end-of-year equity book value, *BVE*.[9] We then form ten portfolios whereby the first (tenth) portfolio contains those observations with the smallest (largest) fraction of *DS* or *RDS*. Second, within each of the ten *DS* or *RDS* portfolios we rank firms according to equity market capitalization and assign each firm to one of three equal-sized groups of firms comprising the small, medium, and large firms. This procedure results in there being ten portfolios within each of the three firm-size groups. It also ensures that the magnitude of *DS* or *RDS* does not vary systematically across the three firm-size groups, and thereby helps to mitigate the confounding effect of firm size when conducting our hedge portfolio tests.[10] Third, we then combine observations from all sample years, retaining the firm size designation and *DS* and *RDS* portfolio rankings. This results in there being three firm-size groups, within each of which there are ten *DS* or *RDS* portfolios.[11] Fourth, within each of the three firm-size groups, for each firm in the ten *DS* or *RDS* portfolios, we compute the risk-adjusted return over all sample years. Fifth, we compute the hedge return by deducting the equally weighted mean risk-adjusted return on the portfolio(s) comprising firms we expect to be most over-valued from the return on portfolio(s) comprising firms we expect to be either undervalued or least over-valued.[12]

We predict that over-valuation is most likely to occur for firms whose income is overstated relative to very comprehensive net income, and where the market fails to understand the economic implications of such overstatement. As noted in Section II, we expect these conditions to be more descriptive for *RDS* than *DS*. Recall that *RDS* is generally non-positive because the accounting procedures that give rise to *RDS* arise from equity transactions that generally are recorded at less than market value. Our hedge strategy is therefore long in firms with least negative *RDS* and short in firms with most negative *RDS*. We employ a similar strategy for *DS*, i.e., go long in firms with most positive *DS* and short in firms with most negative *DS*. As noted above, we do not expect this *DS*-based hedge strategy to yield significant positive (or negative) excess returns.

Following Bernard and Thomas (1990), we compute the hedge return for each of the three firm-size groups by going long (short) in the firms in the top three (bottom three) *DS* or *RDS* portfolios. Combining observations in the top three and bottom three *DS* or *RDS* portfolios confers the benefit of mitigating the potential effects of measurement error in the extreme *DS* or *RDS* portfolios. We employ the hedge portfolio tests to complement the tests based on the forecasting and valuation equations. In particular, if the forecasting and valuation equations yield evidence of mispricing, notably undervaluation of *DS* or *RDS*, then the hedge portfolio tests should yield evidence that excess returns can be earned by exploiting such undervaluation. Conversely, if the forecasting and valuation equations yield no evidence of mispricing, then the hedge portfolio tests

---

[9] Untabulated findings based on *DS* and *RDS* deflated by total assets result in no changes in inferences.

[10] By design, this procedure is a double-conditional sort of first *RDS* (or *DS*) then size. As a consequence, this procedure can fail to adequately control for size differences between long and short *RDS* portfolios. To assess the sensitivity of hedge returns, we reversed the sorting procedure and recomputed hedge returns sorting firm-years using a double-conditional sort of first size then *RDS*. Hedge return findings based on this alternative procedure result in inferences that are substantially the same as those based on the tabulated hedge return findings.

[11] Because firm size is increasing during our sample period, some large firms in early sample years would be considered small firms in later sample years. However, because firm-size groupings are determined annually, our procedure mitigates year effects on our hedge portfolio test inferences.

[12] Untabulated findings based on hedge returns computed with value-weighted portfolio risk-adjusted returns result in no change in inferences. Additional untabulated findings based on cumulative abnormal returns also result in no change in inferences.

should yield evidence that excess returns cannot be earned following our hedge strategy.

Hedge Strategy Implementation Details

To estimate risk-adjusted return, we need a measure of expected stock return. Following Ang and Liu (2004), Ibbotson and Associates (2005), Massa et al. (2005), and Barth et al. (2008), we use the Fama and French (1993) three-factor model, supplemented with the momentum factor (Jegadeesh and Titman 1993; Carhart 1997), with time-varying factor loadings, risk-free rates, and risk premia. We calculate each firm's expected equity return for month t+1 as of month t, $ER_{i,t+1}$, conditional on the expected factor returns in month t+1, based on Equation (6):

$$ER_{i,t+1} = R_{f,t+1} + \beta_{RMRF,i,t+1}(R_{M,t+1} - R_{f,t+1}) + \beta_{SMB,i,t+1}SMB_{t+1} + \beta_{HML,i,t+1}HML_{t+1} + \beta_{MOM,i,t+1}MOM_{t+1}, \quad (6)$$

where $\beta_{RMRF,i,t+1}$, $\beta_{SMB,i,t+1}$, $\beta_{HML,i,t+1}$, and $\beta_{MOM,i,t+1}$ are firm-specific coefficients estimated from Equation (7) below. $R_{M,t+1} - R_{f,t+1}$, SMB_{t+1}, HML_{t+1}, and MOM_{t+1} are the expected monthly Fama-French and momentum factor returns for month t+1. We estimate the expected monthly factor returns for month t by calculating each factor's average monthly return over the 60 months prior to month t. The difference, $R_{M,t} - R_{f,t}$, is the monthly return of the market portfolio in excess of the risk-free rate, HML_t and SMB_t are the monthly returns to the book-to-market and size factor-mimicking portfolios, respectively, as described in Fama and French (1993), and MOM_t is the monthly return to the momentum factor-mimicking portfolio. The risk-adjusted return for firm i in month t+1 is the difference between the firm's realized return in month t+1, $R_{i,t+1}$, and its expected return, $ER_{i,t+1}$. We then use these monthly risk-adjusted returns to compute annual returns. In the hedge return tests, we cumulate return three months after fiscal year-end to ensure that the financial statements are available to the public.

For each firm, we estimate the betas associated with the firm's return to each of the Fama-French and momentum factors by estimating the following monthly time-series regression:

$$R_{i,t} - R_{f,t} = \alpha_i + \beta_{RMRF,i}(R_{M,t} - R_{f,t}) + \beta_{SMB,i}SMB_t + \beta_{HML,i}HML_t + \beta_{MOM,i}MOM_t + \varepsilon_{i,t}, \quad (7)$$

where $R_{i,t} - R_{f,t}$ is the firm's monthly return in excess of the risk-free rate. We estimate Equation (7) using the most recent 60-month returns prior to the month t. This results in estimated coefficients, $\hat{\beta}_{RMTF,i,t}$, $\hat{\beta}_{SMB,i,t}$, $\hat{\beta}_{HML,i,t}$, and $\hat{\beta}_{MOM,i,t}$, which are updated monthly. We define our forecast of each factor beta for month t+1 using the fitted value for that factor for month t, e.g., $\beta_{RMTF,i,t+1} = \hat{\beta}_{RMTF,i,t}$.

Following Doyle et al. (2003), we compute hedge returns over one-, two-, and three-year horizons. We conjecture that if hedge returns continue to increase over longer horizons, then such evidence would be indicative of unmodeled risk differences. Therefore, we expect hedge returns to flatten over the three-year horizon. To avoid imposing the assumption of normality of the distribution of excess returns, we report an additional test for significance of the hedge returns using a t-test based on a boot-strapping procedure. Specifically, we select firm observations that we randomly assign to the ten portfolios. We then calculate the hedge return. We repeat this procedure 1,000 times, thereby generating an empirical distribution that we use to report empirical p-values in addition to conventional t-statistics and their implied p-values.

IV. SAMPLE AND DESCRIPTIVE STATISTICS

We obtain most of the data for estimation of Equation (4) and Equation (5) for 1976–2006 from the Compustat Primary, Secondary, and Tertiary, Full Coverage, and Research Annual Industrial Files. DS_t and RDS_t and are calculated using Compustat and CRSP data as described in the "Computation of *DS* and *RDS*" section. We compute $VCNI_t^a$ as $VCNI_t - rBVE_{t-1}$, where $VCNI_t$

includes both DS_t and RDS_t. Following Barth et al. (1999, 2005), Dechow et al. (1999), Bell et al. (2002), and Landsman et al. (2007), we set r, the cost-of-equity capital, equal to 12 percent, and we require sample firms to have positive equity book value.[13] We also require that sample firms have total assets in excess of $10 million to avoid the undue influence of small firms. To mitigate the effects of outliers, for each variable within each of the three size categories, we treat as missing observations that are in the extreme top and bottom one percentile. For each sample year, firms are ranked according to end-of-year market value of common equity and assigned into one of three equal-sized groups of firms comprising the small, medium, and large firms. We estimate Equation (4) and Equation (5) using unscaled data (Barth and Kallapur 1996). We assess significance of regression coefficients using two-way clustered standard errors, with firm and year clusters (Petersen 2009; Gow et al. 2010).[14] The final sample for estimation of Equation (4) and Equation (5) comprises 37,097 firm-year observations.

We obtain stock return, R, from CRSP and R_f, the one-month Treasury rate, and the Fama-French and momentum factor returns from the Fama-French database (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html). To obtain excess returns per Equation (6), we estimate factor loadings from Equation (7) using monthly return data beginning in 1972.[15] There are 30,383 potential DS firm-year excess return observations. However, because there are 17,579 observations with zero DS, we limit our DS hedge return analysis to the 12,804 nonzero observations. There are 28,346 RDS firm-year excess return observations.

Table 1, Panels A–D present distributional statistics and Panel E presents Pearson and Spearman correlations. Panels A–D reveal that, on average, the market value of equity exceeds the book value of equity for all size firms and mean abnormal earnings, $VCNI^a$, is positive for large firms but negative for medium and small firms. Table 1, Panel E reveals that the explanatory variables in Equation (4) and Equation (5) are correlated with each other, but not so much as to raise collinearity concerns. Although the distributional statistics reported in Panels A–D reveal the variables are skewed, none of the key inferences are affected when the equations are estimated on a per-share basis.

Because typically $P_t^M > P_t^A$ and $N_t > N_{t-1}$, we expect RDS to be negative. Table 1, Panels A–D reveal that RDS turns positive before the 75th percentile. Untabulated statistics reveal that RDS turns positive between the 60th and 70th percentiles for two-thirds of the sample years, and beyond the 70th percentile for the remaining third. Because it is unlikely that equity transactions will give rise to positive RDS, this means that at least some of our observations are measured with error. Assuming this error is unsystematic, the implication of this is a reduction in power of our tests, particularly those relating to the hedge returns.

Untabulated statistics reveal that mean RDS is economically largest (i.e., most negative) for Pharmaceuticals, Services, Food, and Computers, and that mean RDS is economically largest in 1997, 1998, and 2002. To mitigate the impact of particular industries or years overly influencing our results, below we report supplementary findings from hedge return tests that exclude those industries and years with the largest mean RDS values.

---

[13] None of our inferences are affected by assuming alternative values for r, including firm-specific values based on our multi-factor model.

[14] We also compute significance levels using bootstrapping. Untabulated findings result in no changes in inferences from those based on reported findings for RDS. For DS, the forecasting and valuation coefficients are still consistent, but with both significantly positive forecasting and valuation coefficients for small and medium firms.

[15] Although excess returns can be computed through 2006, our sample stops in 2003. This is because we compute hedge returns for one-, two-, and three-year horizons, and to facilitate comparison over returns time, we use a common sample for the full three-year horizon.

TABLE 1

Descriptive Statistics for Equity Market Value, Equity Book Value, Abnormal Earnings, Dirty Surplus, and Really Dirty Surplus for a Sample of 37,097 Firm-Year Observations 1976–2006

Panel A: Distributional Statistics (in $ million)

Variable	Mean	25th %	Median	75th %	Std. Dev.
MVE	2,216.57	89.98	388.66	1,537.13	6,282.58
BVE	894.42	65.67	213.53	745.84	2,151.25
VCNI	102.05	2.01	15.50	73.56	377.09
VCNI[a]	4.43	−13.30	−0.27	12.96	262.01
DS	0.68	0.00	0.00	0.00	49.43
RDS	−18.90	−6.27	−0.42	0.02	150.51

Panel B: Distributional Statistics for Small Firms (in $ million)[a]

Variable	Mean	25th %	Median	75th %	Std. Dev.
MVE	85.88	24.22	50.64	99.54	102.40
BVE	64.65	20.96	40.99	81.83	66.82
VCNI	2.37	−0.18	2.46	6.51	11.83
VCNI[a]	−5.06	−6.79	−1.48	0.78	13.17
DS	0.04	0.00	0.00	0.00	1.23
RDS	−0.73	−0.52	−0.03	0.01	3.76

Panel C: Distributional Statistics for Medium Firms (in $ million)[a]

Variable	Mean	25th %	Median	75th %	Std. Dev.
MVE	531.34	189.03	369.80	700.67	487.35
BVE	283.47	112.81	200.98	366.31	249.43
VCNI	22.95	7.71	18.52	37.65	43.79
VCNI[a]	−7.95	−14.94	0.44	9.11	44.01
DS	0.27	0.00	0.00	0.00	6.62
RDS	−4.61	−4.94	−0.67	0.02	18.51

(continued on next page)

Panel D: Distributional Statistics for Large Firms (in $ million)[a]

Variable	Mean	25th %	Median	75th %	Std. Dev.
MVE	5,914.80	1,240.97	2,685.64	6,020.24	9,741.45
BVE	2,290.81	602.20	1,167.00	2,477.00	3,248.53
VCNI	275.33	46.54	123.25	318.75	608.33
$VCNI^a$	25.62	−52.27	18.65	97.62	446.07
DS	1.70	−2.05	0.00	1.34	84.46
RDS	−50.37	−36.63	−5.32	0.15	254.37

Panel E: Correlations, with Pearson (Spearman) Correlations above (below) the Diagonal

Variable	*MVE*	*BVE*	*VCNI*	$VCNI^a$	*DS*	*RDS*
MVE	1.000	0.802	0.733	0.343	0.044	−0.251
BVE	0.942	1.000	0.761	0.206	0.059	−0.202
VCNI	0.675	0.672	1.000	0.784	0.108	0.167
$VCNI^a$	0.186	0.105	0.628	1.000	0.118	0.431
DS	0.005	0.007	0.062	0.106	1.000	−0.149
RDS	−0.340	−0.287	−0.057	0.125	−0.055	1.000

[a] Firms are ranked for each sample year according to firm size, i.e., equity market value, and assigned into one of three equal-sized groups of firms comprising the small, medium, and large firms.

Variable Definitions:

MVE = market value of common shares outstanding as of fiscal year-end (Compustat #24 ∗ #25);
BVE = book value of common equity as of fiscal year-end (Compustat #60 + #227 − #242);
VCNI = very comprehensive earnings (Compustat #172 − #19 + *DS* + *RDS*);
$VCNI^a$ = abnormal comprehensive earnings, defined as very comprehensive earnings (*VCNI*) minus 0.12 ∗ *BVE* (lagged one year);
DS = dirty surplus, measured as the sum of (1) change in the balance of unrealized gains or losses on marketable securities (change in Compustat #238), (2) change in the cumulative foreign exchange adjustment (change in Compustat #230), and (3) 0.65 times change in additional pension liability in excess of unrecognized prior service costs (change in Min (Compustat #297 – #298, 0)); and
RDS = really dirty surplus, measured as the change in the book value of common equity (Compustat #60 + #227 − #242), less *DS*, less net income (Compustat #172 − #19), plus dividends (Compustat #21), less share price at middle of fiscal year times change in common shares outstanding (Compustat #25, adjusted for stock dividends and splits).

V. RESULTS

Forecasting Equations

Table 2, Panel A presents regression summary statistics from estimating Equation (4). We employ separate estimations for small, medium, and large firms and the pooled sample. Panel A reveals in all cases, the forecasting coefficient for abnormal very comprehensive income, ω_1, is significantly positive. It is also increasing in firm size, which is consistent with greater persistence for larger firms.

The incremental forecasting coefficient for *DS*, ω_2, is significantly different from zero for only the large firms. More importantly, the total *DS* forecasting coefficient, $\omega_1 + \omega_2$, is insignificantly different from zero for all three groups of firms and for the pooled sample (t-statistics = 0.46, 0.98, 0.59, and 0.59).[16] These findings indicate that *DS* is forecasting-irrelevant for $VCNI^a$ for all firms. If investors correctly understand the implications of these persistence findings for valuation, then we should observe valuation irrelevance of *DS* for all firms, i.e., *DS* should have a zero total valuation coefficient in the valuation equation.

The incremental forecasting coefficient for *RDS*, ω_3, is significantly negative in all cases. The total *RDS* forecasting coefficient, $\omega_1 + \omega_3$, is insignificantly different from zero for all size firms and the pooled sample (*t*-statistics = 1.29, 0.84, -0.85, and -0.81). These findings indicate that *RDS* is forecasting-irrelevant for $VCNI^a$ for all firms. As with *DS*, if investors correctly understand the implications of these persistence findings for valuation, then we should observe valuation irrelevance of *RDS* for all firms.

Valuation Equations

Table 2, Panel B reveals the valuation coefficient for $VCNI^a$, α_1, is significantly positive in all cases. It is also increasing in firm size, ranging from 1.05 for small firms to 7.83 for large firms, which is consistent with the pattern of increasing persistence displayed in Table 2, Panel A.

The incremental valuation coefficient for *DS*, α_2, is insignificantly different from zero for small and medium firms and significantly negative for large firms. More importantly, its total coefficient, $\alpha_1 + \alpha_2$, is also insignificantly different from zero for all three groups as well as the pooled sample (t-statistics = 1.03, 1.10, -0.39, and -0.43). This finding is expected based on the findings about the lack of persistence for *DS* revealed in Table 2, Panel A.

The incremental valuation coefficient for *RDS*, α_3, is significantly negative in all cases. Its total coefficient, $\alpha_1 + \alpha_3$, is also significantly negative for all groups (t-statistics = -5.45, -8.37, -4.76, and -4.93). Based on the forecasting coefficient findings in Table 2, Panel A, we expect to observe $\alpha_1 + \alpha_3$ to be insignificantly different from zero for all size firms. However, finding that $\alpha_1 + \alpha_3 < 0$ implies that an incremental dollar of *RDS* increases market value by less than a dollar.[17] Thus investors appear to undervalue the *RDS* component of income, i.e., over-value equity.[18]

---

[16] Throughout we use a 0.05 significance level under a two-sided alternative when evaluating statistical significance.

[17] Note that the market's treatment of *RDS* as having negative persistence can be restated as implying that the market views the benefits to the firm from a transaction that give rise to *RDS* as exceeding its *RDS* costs. For this to occur, the market would have to believe that benefits would flow to the firm in future periods at a level beyond that which could be inferred from the time-series properties of residual income. In other words, *RDS* would play two roles—being both a current period cost and a proxy for an Ohlson-type other information variable. An example of this is the market believing an intangible asset arising from employee stock options is greater than the dilution cost to current shareholders.

[18] The findings in Table 2 could be attributable to variables predictive of future earnings and returns that are correlated with *RDS*. Fairfield et al. (2003) identify two such accounting-based variables: short-term accruals and growth in net operating assets. Untabulated findings reveal that inclusion of these variables in Equation (4) and Equation (5) does not affect any inferences we draw from Table 2.

TABLE 2

Regressions of Abnormal Earnings and Equity Market Value, with Dirty Surplus and Really Dirty Surplus Included as Separate Regressors, for a Sample of 37,097 Firm-Year Observations 1976–2006[a]

Panel A: Summary Statistics from Regression of Abnormal Earnings on Lagged Abnormal Earnings, Dirty Surplus, Really Dirty Surplus, and Equity Book Value

$$VCNI^a_{it+1} = \omega_0 + \omega_1 VCNI^a_{it} + \omega_2 DS_{it} + \omega_3 RDS_{it} + \omega_4 BVE_{it} + \varepsilon_{it+1}$$

		Intercept		*VCNI[a]*		*DS*		*RDS*		*BVE*		Forecasting Relevance Test	
												$\omega_1 + \omega_2 = 0$	$\omega_1 + \omega_3 = 0$
Sample	No. of Obs.	Coeff.	t-stat.	Coeff.	t-stat.	Coeff.	t-stat.	Coeff.	t-stat.	Coeff.	t-stat.	t-stat.	t-stat.
Small[b]	12,257	0.74	2.88*	0.39	15.91*	−0.28	−1.13	−0.35	−13.15*	−0.07	−10.25*	0.46	1.29
Medium[b]	12,213	4.59	3.99*	0.47	17.85*	−0.15	−0.43	−0.45	−11.73*	−0.04	−5.64*	0.98	0.84
Large[b]	12,627	−16.23	−2.64*	0.53	9.94*	−0.43	−2.35*	−0.59	−6.65*	0.00	0.78	0.59	−0.85
Pooled	37,097	−7.98	−4.10*	0.53	10.07*	−0.43	−2.35*	−0.58	−6.65*	0.00	0.54	0.59	−0.81

Panel B: Summary Statistics from Regression of Market Value of Equity on Abnormal Earnings, Dirty Surplus, Really Dirty Surplus, and Equity Book Value

$$MVE_{it} = \alpha_0 + \alpha_1 VCNI^a_{it} + \alpha_2 DS_{it} + \alpha_3 RDS_{it} + \alpha_4 BVE_{it} + u_{it}$$

		Intercept		*VCNI[a]*		*DS*		*RDS*		*BVE*		Valuation Relevance Test	
												$\alpha_1 + \alpha_2 = 0$	$\alpha_1 + \alpha_3 = 0$
Sample	No. of Obs.	Coeff.	t-stat.	Coeff.	t-stat.	Coeff.	t-stat.	Coeff.	t-stat.	Coeff.	t-stat.	t-stat.	t-stat.
Small[b]	12,257	8.85	2.17*	1.05	8.66*	1.21	0.55	−3.56	−6.81*	1.23	12.47*	1.03	−5.45*
Medium[b]	12,213	101.80	3.39*	2.72	8.13*	−0.21	−0.09	−5.09	−9.53*	1.51	13.56*	1.10	−8.37*
Large[b]	12,627	766.66	4.11*	7.83	10.37*	−9.30	−2.33*	−10.88	−9.50*	1.93	15.87*	−0.39	−4.76*
Pooled	37,097	187.08	3.27*	7.76	10.39*	−9.37	−2.34*	−10.96	−9.60*	2.01	16.94*	−0.43	−4.93*

* Indicates coefficient is significantly different from 0 at less than the 0.05 level.

[a] See Table 1 for definitions of all variables.

[b] Firms are ranked for each sample year according to firm size, i.e., equity market value, and assigned into one of three equal-sized groups of firms comprising the small, medium, and large firms.

Dirty Surplus Hedge Returns

Table 3 reports buy-and-hold Fama-French risk-adjusted stock returns for firms in the top and bottom three deciles of firms classified according to the (signed) magnitude of *DS* as a fraction of equity book value at the beginning of the cumulation period. Results are presented separately for small, medium, and large firms, and for the pooled sample. The table presents mean returns for one-, two-, and three-year horizons and the median values of *DS/BVE* for each group, as well as hedge returns and associated t-statistics and empirical p-values.

The findings indicate that for medium and large firms, and for the pooled sample, mean risk-adjusted returns and associated hedge returns for all three horizons are essentially zero. However, for small firms, the mean risk-adjusted returns are positive and monotonically increasing over the three-year horizon, suggesting that Fama-French risk adjustments may not perfectly eliminate the pricing effects of risk. Nonetheless, the small firm hedge returns are zero, indicating that the mismeasurement of expected return for small firms is unrelated to assignment of observations to *DS* portfolios. The small firm hedge returns being zero is consistent with investors pricing *DS* correctly.

Taken together, the findings in Tables 2 and 3 suggest little evidence of mispricing of stocks based on the magnitude of *DS*. Alternatively, the findings suggest investors make no adjustments to reflect *DS* items, but this has no pricing implications because when earnings items are transitory they should only be impounded into price as a result of being included in the book value of equity.

Really Dirty Surplus Hedge Returns

Table 4 reports buy-and-hold Fama-French risk-adjusted stock returns for firms in the top and bottom three deciles of firms classified according to the (signed) magnitude of *RDS* as a fraction of equity book value at the beginning of the cumulation period. Table 4 presents analogous statistics to those presented in Table 3, but are based on *RDS* rather than *DS*. Results are presented separately for small, medium, and large firms, and for the pooled sample.

In contrast to Table 3, the findings in Table 4 generally indicate that risk-adjusted returns differ from zero for all firm groups across all three horizons. In addition, the returns move increasingly away from zero in absolute terms over the investing horizon for almost every portfolio. This is particularly pronounced for small firms, for which the one-, two-, and three-year risk-adjusted returns for the top 30 percent (bottom 30 percent) portfolios are 0.08, 0.12, and 0.17 (0.03, 0.06, 0.08). Recall that the excess returns for small firms in Table 3 also are positive and almost identical for firms in the bottom and top 30 percent portfolios, which we attribute to the difficulty of measuring expected return for small firms. We can therefore treat the excess returns for small firms reported in Table 3 as a benchmark for the measurement error in expected return. Using this approach, we can deduct the average of the two portfolio returns for each investing horizon to arrive at a better estimate of risk-adjusted returns for the small firms. This results in one-, two-, and three-year risk-adjusted returns for the top 30 percent (bottom 30 percent) portfolios of 0.01, 0.02, and 0.02 ($-0.06, -0.04, -0.07$). Note that these additional risk adjustments for the small firms have no effect on their hedge returns because the same adjustment is made to both the long and short positions.[19]

Turning to the hedge returns, Table 4 reveals that they are significantly positive in all three firm-size groups and for the pooled sample. Table 4 also reveals that the hedge returns are

[19] A similar adjustment to excess returns for medium and large firms could be made. However, Table 3 indicates that excess returns for medium and large firms are bounded between -0.01 and 0.02, which suggests that measurement error in expected returns is relatively immaterial for these groups of firms.

TABLE 3

Mean Fama-French Risk-Adjusted Stock Returns in Top and Bottom Three Deciles of Non-Zero Dirty Surplus Deflated by Book Value of Owner Equity for a Sample of 12,804 Firm-Year Observations 1976–2003[a,b]

	No. of Obs.	*MVE*	*DS/BVE*	1-Year	2-Year	3-Year
Small Firms Portfolio[c]						
Bottom 30%	1,314	162.30	−0.022	0.08	0.11	0.16
Top 30%	1,303	146.09	0.016	0.07	0.10	0.15
Hedge Return[d]				−0.01	0.00	0.00
t-stat.				−0.51	−0.13	−0.07
Empirical p-value[e]				0.72	0.56	0.54
Medium Firms Portfolio[c]						
Bottom 30%	1,279	1,075.51	−0.021	−0.01	0.00	0.01
Top 30%	1,272	934.83	0.016	0.02	0.01	0.00
Hedge Return[d]				0.03	0.00	−0.02
t-stat.				1.71^	0.20	−0.62
Empirical p-value[e]				0.03	0.39	0.73
Large Firms Portfolio[c]						
Bottom 30%	1,260	14,073.49	−0.023	0.00	0.00	0.00
Top 30%	1,254	12,525.09	0.016	0.01	−0.01	−0.01
Hedge Return[d]				0.00	−0.01	−0.01
t-stat.				0.34	−0.46	−0.55
Empirical p-value[e]				0.36	0.71	0.73
Pooled Sample Portfolio[c]						
Bottom 30%	3,853	5,014.65	−0.022	0.03	0.04	0.06
Top 30%	3,829	4,462.24	0.016	0.03	0.03	0.05
Hedge Return[d]				0.01	−0.00	−0.01
t-stat.				0.55	−0.19	−0.49
Empirical p-value[e]				0.33	0.58	0.67

^ Indicates hedge return is significantly different from zero at less than the 0.10 level.

[a] See Table 1 for definitions of all variables.

[b] Fama-French risk-adjusted return is a firm's actual return in excess of the risk-free rate less the firm's predicted return based on the Fama-French factor and momentum factor mimicking portfolios, i.e., excess market return, size, book-to-market, and momentum factor.

[c] Firm's size designation and *DS* portfolio ranking are assigned in the following procedure: for each sample year, firms are ranked according to *DS* as a fraction of end of year equity book value, *BVE*, and assigned into ten equal-sized portfolios whereby the first (tenth) portfolio contains those observations with the smallest (largest) fraction of *DS*; within each of the ten *DS* portfolios, firms are ranked according to firm size, i.e., equity market value, and are assigned to one of three equal-sized groups of firms comprising the small, medium, and large firms.

[d] The hedge return is computed by deducting the mean risk-adjusted return on the bottom three deciles portfolio from that on the top three deciles portfolio. The strategy implementation begins three months subsequent to the firm's fiscal year-end.

[e] The proportion of the hedge returns from 1,000 simulations exceeds the observed *DS*-based hedge return. In a simulation, each firm is assigned a random number as the substitute for *DS*, and accordingly the portfolio ranking and size designation following the procedure in table note c.

TABLE 4

Mean Fama-French Risk-Adjusted Stock Returns in Top and Bottom Three Deciles of Really Dirty Surplus Deflated by Book Value of Owner Equity for a Sample of 28,346 Firm-Year Observations 1976–2003[a,b]

	No. of Obs.	*MVE*	*RDS/BVE*	1-Year	2-Year	3-Year
Small Firms Portfolio[c]						
Bottom 30%	2,870	114.23	−0.046	0.03	0.06	0.08
Top 30%	2,860	54.91	0.004	0.08	0.12	0.17
Hedge Return[d]				0.06	0.07	0.09
t-stat.				3.61*	2.61*	2.81*
Empirical p-value[e]				<0.01	0.01	<0.01
Medium Firms Portfolio[c]						
Bottom 30%	2,834	670.28	−0.046	−0.02	−0.05	−0.07
Top 30%	2,824	351.69	0.004	0.01	0.01	0.00
Hedge Return[d]				0.03	0.06	0.07
t-stat.				2.71*	3.92*	3.94*
Empirical p-value[e]				<0.01	<0.01	<0.01
Large Firms Portfolio[c]						
Bottom 30%	2,812	10,156.35	−0.047	−0.02	−0.04	−0.06
Top 30%	2,805	5,625.89	0.004	0.01	0.02	0.02
Hedge Return[d]				0.04	0.06	0.08
t-stat.				4.25*	5.39*	5.83*
Empirical p-value[e]				<0.01	<0.01	<0.01
Pooled Sample Portfolio[c]						
Bottom 30%	8,516	3,615.20	−0.046	−0.01	−0.01	−0.02
Top 30%	8,489	1,994.44	0.004	0.04	0.05	0.07
Hedge Return[d]				0.04	0.06	0.08
t-stat.				5.84*	5.81*	6.00*
Empirical p-value[e]				<0.01	<0.01	<0.01

* Indicates hedge return is significantly different from zero at less than the 0.05 level.

[a] See Table 1 for definitions of all variables.

[b] Fama-French risk-adjusted return is a firm's actual return in excess of the risk-free rate less the firm's predicted return based on the Fama-French factor and momentum factor mimicking portfolios, i.e., excess market return, size, book-to-market, and momentum factor.

[c] Firm's size designation and *RDS* portfolio ranking are assigned in the following procedure: for each sample year, firms are ranked according to *RDS* as a fraction of end of year equity book value, *BVE*, and assigned into ten equal-sized portfolios whereby the first (tenth) portfolio contains those observations with the smallest (largest) fraction of *RDS*; within each of the ten *RDS* portfolios, firms are ranked according to firm size, i.e., equity market value, and are assigned to one of three equal-sized groups of firms comprising the small, medium, and large firms.

[d] The hedge return is computed by deducting the mean risk-adjusted return on the bottom three deciles portfolio from that on the top three deciles portfolio. The strategy implementation begins three months subsequent to the firm's fiscal year-end.

[e] The proportion of the hedge returns from 1,000 simulations exceeds the observed *RDS*-based hedge return. In a simulation, each firm is assigned a random number as the substitute for *RDS*, and accordingly the portfolio ranking and size designation following the procedure in table note c.

increasing over time. For example, for small firms, the one-, two-, and three-year hedge returns are 0.06, 0.07, and 0.09. It is possible that the increasing hedge returns over time could be attributable to an unmodeled risk factor.[20,21]

Additional *RDS* Hedge Return Tests

In this section, we consider several additional tests to examine the sensitivity of the *RDS* hedge returns to previously documented pricing anomalies and risk factors, and to alternative procedures for computing those returns. We also attempt to determine the extent to which different source components of *RDS* account for our hedge return results.

First, we investigate whether the persistence of the *RDS* hedge returns could reflect the effects of pricing anomalies documented in prior research. We consider each anomaly, in turn, by first placing observations into one of ten portfolios based on the magnitude of each anomaly factor. Then, within each portfolio, we rank observations according to the magnitude of *RDS* as a fraction of *BVE*, assigning each observation within each anomaly portfolio to one of ten *RDS* portfolios. This double-sorting process helps to ensure that our findings are not the result of the particular anomaly used in the initial sort.

The pricing anomalies we consider are: the short-term accruals anomaly (Sloan 1996; Xie 2001); the growth in long-term asset accrual anomaly (Fairfield et al. 2003); the long-term pricing reversal anomaly (Daniel and Titman 2006; Fama and French 2008a), and the share repurchase and issuance anomaly (Ikenberry et al. 1995; Daniel and Titman 2006; Mitchell and Stafford 2000; Fama and French 2008a). The share repurchase and issuance anomaly is potentially most closely related to the *RDS* pricing anomaly we document because the latter can only arise from equity transactions. Untabulated findings reveal that *RDS* hedge returns remain significantly positive after controlling for the potential confounding effects of each of the four anomalies. Thus, mispricing associated with previously documented anomalies appears not to account for our finding investor mispricing of *RDS*.

Second, we consider an alternative to our buy-and-hold procedure for computing *RDS* hedge returns. Following Mashruwala et al. (2006), we estimate hedge returns using a monthly calendar-time portfolio approach (Fama and French 1993). Under this approach, hedge returns are calculated based on Jensen's alphas from monthly time-series regressions of hedge portfolio excess returns on Fama-French and momentum factor returns. For each sample year, we assign firms into decile portfolios based on *RDS* as a fraction of equity book value and compute the mean monthly portfolio return for firms in the top and bottom three deciles. We then estimate the following monthly time-series regression for both the high and low *RDS* portfolios:

$$R_{p,t} - R_{f,t} = \alpha_p + \beta_p(R_{M,t} - R_{f,t}) + s_p SMB_t + h_p HML_t + m_p MOM_t + \varepsilon_{p,t}. \tag{8}$$

The resulting Jensen's alpha, α_p, measures the mean monthly return for portfolio p not attributable to the Fama-French and momentum factor returns. We predict that α_p for the high *RDS* portfolio is larger than that for the low *RDS* portfolio. We formally test this by estimating Equation (8) for a hedge portfolio, which is constructed as the difference in mean monthly excess returns for the

---

[20] One possible candidate is firm size, as Table 4 indicates that firms in the bottom 30 percent *RDS* portfolios are roughly twice as large as those in the top 30 percent. To assess the importance of firm size on hedge returns, we regressed excess return on firm size and an indicator variable for whether a firm-year observation is in the top 30 percent *RDS* portfolio. Untabulated findings indicate that the indicator variable coefficient is significantly positive in all cases and over all horizons, and the coefficient on size is insignificant in all specifications.

[21] In addition, it is possible that significant hedge returns are induced by our implementation of risk-adjusting returns. Untabulated hedge returns computed without explicit adjustment for risk yield inferences are consistent with those based on the tabulated findings.

high *RDS* and low *RDS* portfolios,[22] and then testing whether the resulting Jensen's alpha is significantly positive.

Consistent with our predictions, untabulated findings indicate that Jensen's alpha for the high *RDS* portfolio is larger than that for the low portfolio for all three firm-size groups: 0.005 versus 0.003 (small firms); 0.001 versus 0.000 (medium firms); 0.001 versus 0.000 (large firms). For the pooled sample, the high and low portfolio Jensen's alpha are 0.002 and 0.000, respectively. The Jensen's alpha of the hedge portfolio is statistically significant for only the small firms and the pooled sample, 0.003 (t-statistic = 2.25; p-value = 0.03) and 0.002 (t-statistic = 2.27; p-value = 0.02), which in annual terms indicate excess returns to the *RDS*-based hedging strategy of 3.7 percent and 2.4 percent, respectively. These findings contrast with those from the buy-and-hold hedge returns in Table 4, which indicate hedge returns are positive for all three firm-size groups. Ascertaining which approach yields the more reliable results is not straightforward. For example, the buy-and-hold approach has the advantage of updating the individual stock's expected return on a monthly basis using out-of-sample estimation, and the alpha approach assumes that factor betas are constant during the test period.[23]

Third, we attempt to determine the extent to which different source components of *RDS* account for our hedge return results. Findings from the forecasting and valuation tests are consistent with, in the aggregate, *RDS* components of very comprehensive net income being transitory, but investors failing to understand this and over-valuing equity. Ideally, we would like to identify separate components of *RDS*, determine their persistence, and then ascertain which components investors appear to fail to price correctly. Because we face the same data limitations as investors, we can only do this indirectly by sequentially excluding and including firm-years in which *RDS* is more likely to be attributable to one particular type of transaction. First, to focus on the potential pricing effects of pooling-of-interests transactions, we recomputed hedge returns (1) excluding firm-years with corporate mergers or acquisitions, and (2) using only firm-years with corporate mergers.[24] Second, to focus on the potential pricing effects of employee stock options, we recomputed hedge returns using (1) only observations beginning in 1995, and (2) only observations before 1995. Because SFAS No. 123 required disclosure of weighted average exercise price for employee stock options after 1995, hedge returns might be expected to fall or even disappear in this latter period. Third, to focus more generally on the pricing effects of dilutive transactions, including warrants, convertible instruments, as well as stock options, we recomputed hedge returns for portfolios sorting firm-year observations on the difference between basic and diluted earnings per share as a fraction of equity book value. Finally, because *RDS* is concentrated in particular industries and years, as noted in Section IV, we sequentially calculated hedge returns excluding the six industries and the five years with the most negative *RDS*.

Untabulated findings from each of these additional analyses reveal that hedge returns remain significant in all cases. These findings are consistent with our tests lacking power to trace the precise sources of *RDS*. They are also consistent with *RDS* being correlated with an unmodeled risk factor.

---

[22] The independent variables (risk factors) are the same for each portfolio and are therefore included in the hedge portfolio without adjustment.

[23] For a discussion of the strengths and weaknesses of the various approaches, see Fama and French (2008b).

[24] Ideally, we would compute hedge returns for subsamples excluding and including only those firm-years with mergers accounted for under pooling-of-interests. However, because such identifying information is only available to us using the Securities Data Corporation beginning in the middle of our sample period, we elected to cast the net wider to take account of mergers and acquisitions during our sample period.

VI. CONCLUSIONS

The question we address in this study is whether firms' share prices correctly reflect two accounting measures, dirty surplus and really dirty surplus. We find that both dirty surplus and really dirty surplus are forecasting-irrelevant for abnormal very comprehensive income for all firm-size groups. Taking these results at face value (i.e., assuming that the forecasting and valuation equations correctly capture the time-series properties of these variables), if investors correctly understand the implications of these persistence findings for valuation, then each kind of dirty surplus should be valuation-irrelevant for all firms. This prediction is borne out in the case of dirty surplus. In contrast, the findings indicate that investors appear to undervalue really dirty surplus, which is consistent with the premise that investors are unable to assess the economic implications of really dirty surplus transactions. However, the possibility remains that the mismatch of the really dirty surplus forecasting and valuation coefficients is the result of model misspecification rather than mispricing.

Our buy-and-hold hedge return results support the findings from the tests linking the forecasting and valuation equations. Hedge returns are insignificantly different from zero when based upon dirty surplus, regardless of firm size and investing horizon. In contrast, buy-and-hold hedge returns based on really dirty surplus are significantly positive for all three firm-size groups as well as for the pooled sample. Findings from additional tests reveal that inferences relating to hedge returns are insensitive to an alternative procedure to measuring hedge returns, to including controls for four previously identified mispricing anomalies, and to sampling procedures designed to focus on sources of really dirty surplus.

Taken together, the findings are consistent with investors failing to understand the lack of persistence of really dirty surplus, and therefore apparently over-valuing firms that have large negative really dirty surplus. However, because we are unable to trace the sources of really dirty surplus to particular types of equity transactions, we cannot determine the extent to which potential mispricing arises from each type of transaction. However, even if we could trace the sources of really dirty surplus, any resulting hedge returns might still be attributable to an unmodeled risk factor.

REFERENCES

Ang, A., and J. Liu. 2004. How to discount cash flows with time-varying expected returns. *The Journal of Finance* 59 (6): 2745–2783.

Barth, M. E., and S. Kallapur. 1996. Effects of cross-sectional scale differences on regression results in empirical accounting research. *Contemporary Accounting Research* 13 (2): 527–567.

———, W. H. Beaver, J. M. Hand, and W. R. Landsman. 1999. Accruals, cash flows, and equity values. *Review of Accounting Studies* 4 (3–4): 205–229.

———, ———, ———, and ———. 2005. Accruals, accounting-based valuation models, and the prediction of equity values. *Journal of Accounting, Auditing & Finance* 20 (4): 311–345.

———, Y. Konchitchki, and W. R. Landsman. 2008. Cost of capital and financial statements transparency. Working paper, Stanford University.

Bell, T., W. R. Landsman, B. Miller, and S. Yeh. 2002. The valuation implications of employee stock option accounting for profitable computer software firms. *The Accounting Review* 77 (4): 971–996.

Bernard, V., and J. Thomas. 1989. Post-earnings-announcement drift: Delayed price response or risk premium? *Journal of Accounting Research* 27 (Supplement): 1–48.

———, and ———. 1990. Evidence that stock prices do not fully reflect the implications of current earnings for future earnings. *Journal of Accounting and Economics* 13 (4): 305–340.

Biddle, G., and J. Choi. 2006. Is comprehensive income useful? *Journal of Contemporary Accounting and Economics* 2 (1): 1–32.

Bradshaw, M., and R. G. Sloan. 2002. GAAP versus the Street: An empirical assessment of two alternative

definitions of earnings. *Journal of Accounting Research* 40 (1): 41–66.

Brown, L., and K. Sivakumar. 2003. Comparing the value relevance of two operating income measures. *Review of Accounting Studies* 8 (4): 561–572.

Burgstahler, D., J. Jiambalvo, and T. Shevlin. 2002. Do stock prices fully reflect the implications of special items for future earnings? *Journal of Accounting Research* 40 (3): 585–612.

Carhart, M. 1997. On persistence in mutual fund performance. *The Journal of Finance* 52 (1): 57–82.

Chambers, D., T. J. Linsmeier, C. Shakespeare, and T. Sougiannis. 2007. An evaluation of SFAS No. 130 comprehensive income disclosures. *Review of Accounting Studies* 12 (4): 557–593.

Christensen, P., and G. A. Feltham. 2003. *Economics of Accounting*. Boston, MA: Kluwer Academic Publishers.

Core, J., W. Guay, and S. P. Kothari. 2002. The economic dilution of employee stock options: Diluted EPS for valuation and financial reporting. *The Accounting Review* 77 (3): 627–652.

Daniel, K., and S. Titman. 2006. Market reactions to tangible and intangible information. *The Journal of Finance* 61 (4): 1605–1643.

Dechow, P. M., A. P. Hutton, and R. G. Sloan. 1999. An empirical assessment of the residual income valuation model. *Journal of Accounting and Economics* 26 (1–3): 1–34.

Dhaliwal, D., K. R. Subramanyam, and R. Trezevant. 1999. Is comprehensive income superior to net income as a measure of firm performance? *Journal of Accounting and Economics* 26 (1–3): 43–67.

Doyle, J. T., R. J. Lundholm, and M. T. Soliman. 2003. The predictive value of expenses excluded from 'pro forma' earnings. *Review of Accounting Studies* 8 (2–3): 145–174.

Fairfield, P. M., J. S. Whisenant, and T. L. Yohn. 2003. Accrued earnings and growth: Implications for future profitability and market mispricing. *The Accounting Review* 78 (1): 353–371.

Fama, E. F., and K. R. French. 1993. Common-size risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33 (1): 3–56.

———, and ———. 2008a. Average returns, B/M, and share issues. *The Journal of Finance* 63 (6): 2971–2995.

———, and ———. 2008b. Dissecting anomalies. *The Journal of Finance* 63 (4): 1653–1678.

Feltham, G. 1996. Valuation, clean surplus accounting, and anticipated equity transactions. Working paper, University of British Columbia.

———, and J. Ohlson. 1995. Valuation and clean surplus accounting for operating and financial activities. *Contemporary Accounting Research* 11 (2): 689–731.

Gow, I. D., G. Ormazabal, and D. J. Taylor. 2010. Correcting for cross-sectional and time-series dependence in accounting research. *The Accounting Review* 85 (2): 483–512.

Ibbotson, and Associates. 2005. *Cost of Capital 2005 Yearbook*. Chicago, IL: Ibbotson and Associates.

Ikenberry, D., J. Lakonishok, and T. Vermaelen. 1995. Market underreaction to open market share repurchases. *Journal of Financial Economics* 39 (2–3): 181–208.

Jegadeesh, N., and S. Titman. 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance* 48 (1): 65–91.

Landsman, W. R., B. L. Miller, and S. Yeh. 2007. Implications of components of income excluded from pro forma earnings for future profitability and equity valuation. *Journal of Business Finance & Accounting* 34 (3–4): 650–675.

———, K. Peasnell, P. Pope, and S. Yeh. 2006. Which approach to accounting for employee stock options best reflects market pricing? *Review of Accounting Studies* 11 (2–3): 203–245.

Mashruwala, C., S. Rajgopal, and T. Shevlin. 2006. Why is the accrual anomaly not arbitraged away? The role of idiosyncratic risk and transactions costs. *Journal of Accounting and Economics* 42 (1–2): 3–33.

Massa, M., U. Peyer, and Z. Tong. 2005. Limits of arbitrage and corporate financial policy. Working paper, INSEAD.

Mitchell, M. L., and E. Stafford. 2000. Managerial decisions and long-term stock price performance. *The Journal of Business* 73 (3): 287–329.

O'Hanlon, J., and P. Pope. 1999. The value-relevance of U.K. dirty surplus accounting flows. *The British Accounting Review* 31 (4): 459–482.

Ohlson, J. A. 1995. Earnings, equity book values, and dividends in equity valuation. *Contemporary Accounting Research* 11 (2): 661–687.

——. 1999. On transitory earnings. *Review of Accounting Studies* 4 (3–4): 145–162.

Petersen, M. A. 2009. Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies* 22 (1): 435–480.

Sloan, R. 1996. Do stock prices fully reflect information in accruals and cash flows about future earnings? *The Accounting Review* 71 (3): 289–316.

Xie, H. 2001. The mispricing of abnormal accruals. *The Accounting Review* 76 (3): 357–373.

THE ACCOUNTING REVIEW
Vol. 86, No. 1
2011
pp. 259–286

American Accounting Association
DOI: 10.2308/accr.00000009

Can Big 4 versus Non-Big 4 Differences in Audit-Quality Proxies Be Attributed to Client Characteristics?

Alastair Lawrence
Miguel Minutti-Meza
Ping Zhang
University of Toronto

ABSTRACT: This study examines whether differences in proxies for audit quality between Big 4 and non-Big 4 audit firms could be a reflection of their respective clients' characteristics. In our analyses, we use three audit-quality proxies—discretionary accruals, the *ex ante* cost-of-equity capital, and analyst forecast accuracy—and employ propensity-score and attribute-based matching models in attempt to control for differences in client characteristics between the two auditor groups while estimating the audit-quality effects. Using these matching models, we find that the effects of Big 4 auditors are insignificantly different from those of non-Big 4 auditors with respect to the three audit-quality proxies. Our results suggest that differences in these proxies between Big 4 and non-Big 4 auditors largely reflect client characteristics and, more specifically, client size. We caution the reader that this study has not resolved the question, although we hope that it encourages other researchers to explore alternative methodologies that separate client characteristics from audit-quality effects.

Keywords: *Big 4 versus non-Big 4 audit quality; discretionary accruals; ex ante cost-of-equity capital; analyst forecast accuracy; propensity-score matching; attribute-based matching.*

Data Availability: *Data are publicly available from sources identified in the article.*

We have received valuable comments from Steven Kachelmeier (senior editor), Francesco Bova, Gus DeFranco, Ole-Kristian Hope, Stephannie Larocque, Hai Lu, Flora Niu, Gordon Richardson, Kevin Veenstra, Dushyantkumar Vyas, two anonymous reviewers, and workshop participants at the University of Toronto and Wilfrid Laurier University. Professor Zhang acknowledges the financial support provided by the Social Sciences and Humanities Research Council of Canada.

Editor's note: Accepted by Steven Kachelmeier.

Submitted: June 2008
Accepted: June 2010
Published Online: January 2011

I. INTRODUCTION

The auditing literature generally concludes that the audit quality of Big 4 auditors is superior to that of non-Big 4 auditors.[1] DeAngelo (1981) argues that accounting firm size is a proxy for auditor quality, as no single client is important to larger accounting firms and, hence, larger accounting firms are less likely than smaller accounting firms to compromise their independence. Dopuch and Simunic (1980) propose that larger accounting firms provide higher quality services because they have greater reputations to protect. Furthermore, it could be argued that Big 4 firms provide superior audit quality as their sheer size can support more robust training programs, standardized audit methodologies, and more options for appropriate second partner reviews.

However, there are also arguments as to why Big 4 and non-Big 4 firms could provide comparable audit quality. First, Big 4 and non-Big 4 firms are held to the same regulatory and professional standards, and thus both types of audit firms must adhere to a reasonable level of quality. Second, as "non-Big 4 auditors have superior knowledge of local markets and better relation with their clients" (Louis 2005, 77), these factors may enable non-Big 4 firms to better detect irregularities. Of course, the converse argument could be made that closer relationships among non-Big 4 accounting firms and their clients could potentially lead to a compromise of independence; however, the net effect of these counteracting forces is unclear. Third, the inability of non-Big 4 firms to obtain affordable insurance coverage may actually increase the audit effort of non-Big 4 firms relative to Big 4 firms because smaller audit firms cannot obtain a similar level of backing from insurance companies. This notion is supported by the U.S. Government Accountability Office (GAO) report issued in 2008 indicating that non-Big 4 auditors are struggling to obtain affordable liability insurance coverage (GAO 2008, 55). Moreover, the previous GAO audit firm concentration report issued in 2003 finds that "smaller firms lacked the size needed to achieve economies of scale to spread their litigation risk and insurance costs across a larger capital base" (GAO 2003, 49), suggesting that insurance fees are a fixed cost to audit partners. Finally, CPAs frequently switch between Big 4 and non-Big 4 firms, and knowledge transfers can dilute the potential for one type of audit firm to become superior. For example, upon the fallout of Arthur Andersen (AA), Grant Thornton acquired AA's Albuquerque, Charlotte, Columbia, Greensboro, and Orlando offices and separately hired many former middle-market AA audit partners from other offices (Dow Jones News Service 2002). Hence, it is not obvious from theory or intuition that Big 4 firms should be superior to non-Big 4 firms.

Nonetheless, numerous empirical studies, following the theoretical support of DeAngelo (1981) and Dopuch and Simunic (1980), and using a variety of audit-quality proxies, find evidence suggesting that Big 4 auditors provide higher-quality audits than non-Big 4 auditors (e.g., Palmrose 1988; Becker et al. 1998; Khurana and Raman 2004; Behn et al. 2008). Given that the distributions of client characteristics significantly differ across Big 4 and non-Big 4 firms, an important consideration of these prior studies is whether the empirical findings simply reflect client and not accounting firm characteristics. Our research objective is to investigate whether the Big 4 treatment effect may be attributable to client characteristics.

The question of Big 4 superiority is important, given that many studies rely on the Big 4 versus non-Big 4 distinction as an audit-quality proxy. Hence, it is prudent to confirm that this distinction does not simply reflect client characteristics.[2] Furthermore, incorrectly classifying Big

---

[1] We use Big 4 as a generic term encompassing the Big 8, Big 6, Big 5, and Big 4 to reflect the consolidation of these firms.

[2] Several studies outside the auditing literature have used the Big 4 distinction to proxy for a higher level of audit quality; for example, Beatty (1989), Guenther and Willenborg (1999), Mitton (2002), Smart and Zutter (2003), and Gul et al. (2009).

4 auditors as superior to non-Big 4 auditors has unnecessary negative ramifications for smaller auditors, such as audit committee's auditor selection bias (CFA Institute Center 2009) and discriminatory clauses in loan and underwriting agreements (DeAngelo 1981), which could result in a loss of current and future clients.

Using three audit-quality proxies—discretionary accruals, the *ex ante* cost-of-equity capital, and analyst forecast accuracy—we replicate previous findings that document differences in quality between Big 4 and non-Big 4 audit firms. Following Rosenbaum and Rubin (1983), Li and Prabhala (2007), and Francis et al. (2010), we use propensity-score matching models in an attempt to control for differences in client characteristics between the two auditor groups while estimating auditor treatment effects. Moreover, we match on specific client characteristics (attribute-based matching), such as client size and return on assets, to identify whether Big 4 versus non-Big 4 differences in audit-quality proxies can be attributed to specific client characteristics. Using these matching models, we find that the treatment effects of Big 4 auditors are insignificantly different from those of non-Big 4 auditors with respect to our three audit-quality proxies. Furthermore, the attribute-matching results provide evidence for the argument that differences in the foregoing proxies between Big 4 and non-Big 4 clients largely reflect client size. Using our full samples, we control for an extensive list of firm characteristics known to influence audit quality and find that the Big 4 effect is generally insignificant, indirectly supporting the argument that the Big 4 distinction may reflect client and not auditor characteristics.

Our results must be interpreted with due regard to their limitations and to the caveats of matching models, discussed in more detail throughout the article. This study has not resolved the underlying question as to whether differences in audit-quality proxies between Big 4 and non-Big 4 auditors can be attributed to client characteristics, but we hope that it encourages other researchers to explore alternative methodologies that further disentangle client characteristics from audit-quality effects.

II. RELATED STUDIES AND MEASURES OF AUDIT QUALITY

As the main observable outcome of an audit is the standardized audit report, researchers have used various proxies in an attempt to assess audit quality and, in turn, determine whether a differential in audit quality exists. An extensive branch of audit differentiation research focuses on the quality of the client's financial statements, in which discretionary accruals are often used as a proxy for audit quality, as they reflect the auditor's constraint over management's reporting decisions. Becker et al. (1998) find that Big 4 clients report lower absolute discretionary accruals than non-Big 4 clients. Francis et al. (1999) suggest that Big 4 auditors constrain opportunistic and aggressive reporting because their clients have higher total accruals but lower discretionary accruals. Krishnan (2003) finds a greater association between discretionary accruals and future earnings for Big 4 than for non-Big 4 clients. Following this literature, we use discretionary accruals as our first measure of audit quality. The benefit of this measure is that it reflects the auditor's enforcement of accounting standards. However, a weakness is that it only partially captures the effectiveness of an audit in constraining earnings management, as discretionary accruals not only reflect management's opportunism, but also management's signaling attempts and random noise, as noted by Guay et al. (1996).

The valuation literature suggests that Big 4 auditors provide more assurance to the market than non-Big 4 auditors. The underlying intuition for using valuation proxies as a measure of audit quality is due to the fact that if market participants perceive that the Big 4 clients have more credible earnings than those of the non-Big 4 clients then, *ceteris paribus*, the Big 4 clients should receive a break in their cost-of-equity capital. For example, Khurana and Raman (2004) find that Big 4 clients have a lower *ex ante* cost of capital than non-Big 4 clients in the U.S.; however, they do not find such a difference in Australia, Canada, or Great Britain. Following Khurana and

Raman (2004), we use the *ex ante* cost-of-equity capital as our second audit-quality proxy to capture the capital market's perception of the financial reporting credibility of Big 4 and non-Big 4 clients.

More recently, Behn et al. (2008) include analyst forecast accuracy as an audit-quality proxy. They argue that if one type of auditor increases the reporting reliability of earnings in comparison to the other type, then, *ceteris paribus*, analysts of the superior type's clients should be able to make more accurate forecasts of future earnings than those analysts of the non-superior type's clients. Using this reasoning, Behn et al. (2008) find that analysts of Big 4 clients have higher forecast accuracy than analysts of non-Big 4 clients. We use analyst forecast accuracy as our third audit-quality measure to proxy for an enhanced level of decision making by sophisticated financial statement users.

In summary, we employ three audit-quality proxies—discretionary accruals, the *ex ante* cost-of-equity capital, and analyst forecast accuracy—to capture various aspects of audit quality in comparing Big 4 to non-Big 4 firms.[3]

III. MATCHING PROCEDURES AND TESTS OF NONLINEARITY

We use propensity-score matching models, developed by Rosenbaum and Rubin (1983), to match on a broad range of client characteristics and use attribute-based matching to examine whether the Big 4 distinction can be attributed to specific client characteristics. Propensity-score matching models match observations based on the probability of undergoing the treatment, which in our case is the probability of selecting a Big 4 auditor. Matching models have important features appropriate for our setting. First, these models generate samples in which the clients of Big 4 and non-Big 4 auditors are similar, providing a natural framework to parse out the effects of auditor and client characteristics on the audit-quality proxy. Second, previous studies examining the existence of a differential in audit quality typically use Heckman (1979) selection models that rely on a specific functional form to provide an indirect estimate of auditor treatment effects. Matching models do not rely on a specific functional form and provide a more direct estimate of the treatment effects (Li and Prabhala 2007). Moreover, selection models pertaining to auditor choice could estimate treatment effects with error by failing to meet the exclusion restrictions because it is difficult to identify variables that influence the auditor selection (first-stage regression) and not the audit-quality proxy (second-stage regression). Last, matching models mitigate the potential impact of nonlinearities in estimating the treatment effects when the underlying functional form is nonlinear. A recent example of a study using a propensity-score matching model in the accounting literature is Armstrong et al. (2010), who find that the previously documented relation between equity-based compensation and accounting irregularities does not hold using propensity-score matching models.

Notwithstanding their benefits, matching models have caveats. First, matching models rely on the assumption that the effects of unobservables are not pertinent to estimating treatment effects. Second, matching results in using subsamples of the population, and hence, as noted by Cram et al. (2009), matching reflects a trade-off between identifying the treatment effects and generalizing the results to the full population. Third, matching results in a different auditor composition in the

---

[3] Other branches of the auditor differentiation literature use litigation, going concern, and fraud frequencies as audit-quality proxies. These branches suggest that Big 4 firms are sued and sanctioned less (Palmrose 1988; Feroz et al. 1991), have lower thresholds for and greater accuracy in issuing opinions (Francis and Krishnan 1999; Lennox 1999; Weber and Willenborg 2003; Geiger and Rama 2006), and have a lower incidence of fraud among their clients than non-Big 4 firms (Farber 2005). We do not use these proxies because the infrequent occurrence of these events results in small sample sizes with low power, potentially biasing in favor of our inferences. For example, Dechow et al. (1996) and Beneish (1999), using small samples, do not find a significant difference between the fraud incidence of Big 4 and non-Big 4 clients.

matched samples than in the full population. These features of matching models could bias our findings due to the reduced power of the tests or due to systematic differences in the subsamples from the full population. Fourth, the empirical inability to match on pre-treatment attributes and control for alternative treatments could result in a bias if the matching variables are affected by the auditor choice. Despite our attempts, we cannot fully rule out concerns that our inferences are affected by *ex post* matching or alternative treatments. In the "Sensitivity Analyses" section, we document a series of robustness tests in attempt to mitigate some of these concerns.

We use a logit model to estimate the probability of selecting a Big 4 auditor, as it is the most prevailing approach for estimating propensity scores (Guo and Fraser 2010, 135).[4] Given that matching models do not require exclusion restrictions, the general rule is to include a comprehensive list of attributes when estimating the propensity score (Li and Prabhala 2007). We estimate the propensity score for each audit-quality proxy analysis, including variables from the selection model used in Chaney et al. (2004) and the respective audit quality regression, as follows:[5]

$$BIG4_{i,t} = \beta_0 + \beta_1 LOG\_ASSETS_{i,t} + \beta_2 ATURN_{i,t} + \beta_3 CURR_{i,t} + \beta_4 LEV_{i,t} + \beta_5 ROA_{i,t} + \Sigma PROXY\_CONTROLS_{i,t} + Industry\_FE + Year\_FE + \varepsilon_{i,t}, \quad (1)$$

where for firm i and fiscal year t:

$BIG4$ = 1 if the client has a Big 4 auditor in year$_t$, and 0 otherwise;
LOG_ASSETS = natural logarithm of total assets at the end of year$_t$;
$ATURN$ = sales$_t$/total assets$_{t-1}$;
$CURR$ = current assets$_t$/current liabilities$_t$;
LEV = (long term debt$_t$ plus debt in current liabilities$_t$)/average total assets$_{t-1}$;
ROA = net income$_t$/average total assets$_{t-1}$; and,
$PROXY\_CONTROLS$ = control variables used and described in the respective audit-quality analysis.

We estimate the propensity-score model by including all audit proxy control variables for the respective analysis, but we exclude variables from the regression if they are redundant to the first five variables in predicting auditor choice.[6] We then match, without replacement, a non-Big 4 audit client with a Big 4 audit client that has the closest predicted value from Equation (1) within a maximum distance of 3 percent.[7] Using this caliper distance, we match approximately 80 percent, 99 percent, and 97 percent of non-Big 4 audit clients to Big 4 audit clients for the discretionary accruals, the *ex ante* cost-of-equity capital, and the analyst forecast-accuracy samples, respectively. In effect, this procedure creates a pseudo "random" sample in which the auditor type is randomly allocated to both the treatment and control groups (Heckman and Navarro-Lozano 2004), such that any resulting differences between the two groups should reflect the treatment

[4] All findings documented in this study are robust to using a probit model instead of a logit model to calculate propensity scores. In addition, we find that the logit models result in more balanced matched samples than using the probit models.

[5] We also estimate and match the propensity scores within each industry and year rather than using indicator variables. We find that all results using this alternative specification are similar to the tabulated results.

[6] Alternatively, we estimate the propensity-score model by including all the variables in the model as specified above even when a proxy control variable is similar to a variable in the auditor selection model. We find that the results are robust to including all the redundant variables simultaneously or to including only one redundant variable at a time.

[7] Inferences are the same whether we match with or without replacement. A caliper is the distance between the predicted probabilities of choosing the treatment between matched observations (Dehejia and Wahba 2002).

effect and not pre-existing client characteristics (Heckman et al. 1997, 1998; Dehejia and Wahba 1999, 2002). Hence, differences in means between the treatment and control groups should be sufficient to estimate the treatment effects (Dehejia and Wahba 2002; Dehejia 2005). Nonetheless, we also use multivariate analyses to further control for any remaining characteristic imbalances between the two client groups as well as general cross-sectional characteristic variations.

We employ the Ramsey Regression Equation Specification Error Test (Ramsey 1969; hereafter, RESET test), to identify specific client characteristics that are nonlinear to both the auditor choice and the audit-quality proxy.[8] For each propensity-score regression and each audit-quality regression, we add squared and cubed versions of all control variables, as described in Sections V–VII. Using an F-test, we test the null hypothesis that all of the nonlinear terms of each control variable equal zero. If significant nonlinearities exist in the data and the Big 4 partitioning variable is correlated with these nonlinear variables, then these conditions could potentially confound the inferences of a standard OLS regression. For the attribute-based matching, we create single-variable matched samples in each analysis, using the propensity-score methodology controlling for year and industry effects, based on those variables that are nonlinear to both the auditor choice and the audit-quality proxy.

IV. SAMPLE SELECTION

For our analyses, we use firm-year data from 1988 to 2006. We restrict our analyses to this time period because reported operating cash flows as per SFAS No. 95 (FASB 1987) are only available starting from 1988. For our discretionary accruals sample, we use data from Compustat. After deleting firms in the financial services industries (SIC codes between 6000–6999) and after imposing all the necessary requirements to calculate the discretionary accruals regression variables, we obtain a sample of 72,600 firm-year observations, in which 59,323 and 13,277 are Big 4 and non-Big 4 clients, respectively.[9] Our *ex ante* cost-of-equity sample reflects the intersection of Compustat, I/B/E/S Summary, and CRSP data. After imposing the necessary requirements to calculate the *ex ante* cost-of-capital regression variables, we obtain a sample of 25,068 firm-year observations, in which 23,856 and 1,212 reflect Big 4 and non-Big 4 clients, respectively.[10] Similar to the *ex ante* cost-of-equity sample, the forecast-accuracy sample reflects the intersection of Compustat, I/B/E/S, and CRSP data; however, the accuracy sample requires individual analyst forecast data from the I/B/E/S Detail files, whereas the *ex ante* cost-of-equity sample requires consensus forecast data from the I/B/E/S Summary files. After imposing the necessary requirements to calculate the forecast-accuracy regression variables, we obtain a sample of 28,037 firm-year observations, in which 26,521 and 1,516 reflect Big 4 and non-Big 4 clients, respectively.

---

[8] As in predicting the propensity scores, we exclude variables from the RESET analysis if the two regressions contain redundant variables. The comparison of the RESET findings across the three audit-quality proxies is inherently limited given that we adopt the audit-quality regression models for each proxy from previous studies; thus, there are limited control variables that are common to all three analyses.

[9] We delete firms with negative total assets, sales, debt, and market values of equity, as such observations introduce noise into the discretionary accruals regressions; moreover, raw variables from Compustat are winsorized at the 1 percent and 99 percent levels. We find that inferences for our three analyses hold without winsorizing except for those pertaining to the propensity-score matched sample in the analyst forecast accuracy analysis, as significant outliers reduce the effectiveness of the multivariate propensity-score matching.

[10] For the *ex ante* cost-of-equity, and forecast-accuracy samples, we delete firms in the financial services industries (SIC codes between 6000–6999) and all variables are winsorized at the 1 percent and 99 percent levels.

V. ANALYSIS 1: DISCRETIONARY ACCRUALS

Method

We measure performance-adjusted discretionary accruals using the modified Jones model as recommended by Kothari et al. (2005; hereafter KLW).[11] To test whether the Big 4 versus non-Big 4 discretionary accruals differences could be attributed to client characteristics, we use the following model:

$$ADA_{i,t} = \beta_0 + \beta_1 BIG4_{i,t} + \beta_2 LOG\_MKT_{i,t} + \beta_3 ROA_{i,t-1} + \beta_4 LEV_{i,t-1} + \beta_5 CURR_{i,t-1} + Industry\_FE + Year\_FE + \varepsilon_{i,t}, \quad (2)$$

where for firm i and fiscal year t:

ADA = absolute discretionary accruals as per KLW in $year_t$;
$BIG4$ = 1 if the client has a Big 4 auditor in $year_t$, and 0 otherwise;
LOG_MKT = natural logarithm of market value of equity at the end of $year_t$;
ROA = income before extraordinary $items_{t-1}$/average total $assets_{t-1}$;
LEV = (long-term $debt_{t-1}$ plus debt in current $liabilities_{t-1}$)/average total $assets_{t-1}$; and
$CURR$ = current $assets_{t-1}$/current $liabilities_{t-1}$.

BIG4 is our main variable of interest and is used in a manner consistent with prior research. We include the variable *LOG_MKT* to control for client size. Following Butler et al. (2004), we include *ROA*, as the KLW procedure does not eliminate all the variation in accruals due to firm-specific performance. Also in line with Butler et al. (2004), we include *LEV* and *CURR* to control for the impact of financial risk on discretionary accruals.

Results

Table 1 presents the descriptive statistics for both the full and propensity-score matched samples. There are 72,600 firm-year observations in the full sample, of which 59,323 (77.6 percent) and 13,277 (22.4 percent) reflect Big 4 and non-Big 4 clients, respectively. The descriptive statistics for the full sample indicate that Big 4 and non-Big 4 auditors have significantly different clienteles. Big 4 clients are approximately ten times the size of non-Big 4 clients: the mean market value and total assets of Big 4 clients are \$2.2 billion and \$2.0 billion, respectively, whereas the mean market value and total assets of non-Big 4 auditors are \$234 million and \$258 million, respectively. Moreover, Big 4 clients are more profitable and leveraged than are non-Big 4 clients, and have significantly less discretionary accruals and current assets than non-Big 4 clients.

Using Equation (1) to calculate the propensity scores and imposing a caliper distance of 3 percent, we obtain a propensity-score matched sample of 21,130 firm-years, of which 10,565 are Big 4 clients and 10,565 are non-Big 4 clients. The propensity-score model appears effective in forming a balanced sample of Big 4 and non-Big 4 clients, as all accrual control variables in the

---

[11] The KLW model is as follows and is estimated by year and by two-digit SIC code, scaling by average lagged assets:

$$TOTAL\_ACCRUALS_{i,t} = \beta_0 + \beta_1(1/ASSETS_{i,t}) + \beta_2(\Delta SALES_{i,t} - \Delta REC_{i,t}) + \beta_3 PPE_{i,t} + \varepsilon_{i,t},$$

where for firm i and fiscal year t, *TOTAL_ACCRUALS* equals (net income before extraordinary items minus operating cash flows from continuing $operations_t$)/average total $assets_{t-1}$; *ASSETS* equals average total $assets_{t-1}$; ΔREC equals (change in accounts receivable from $year_{t-1}$ to $year_t$)/average total $assets_{t-1}$; *PPE* equals net property, plant, and equipment in $year_t$/average total $assets_{t-1}$; and ε equals the pre-KLW estimated discretionary accrual. From each firm's estimated pre-KLW discretionary accrual, we subtract the estimated pre-KLW discretionary accrual of the closest ROA firm in the same two-digit SIC code. The resulting normalized error term is the KLW discretionary accrual measure.

TABLE 1

Discretionary Accruals Analysis: Descriptive Statistics Full and Propensity-Score Matched Samples

	Full Sample				Propensity-Score Matched Sample: Matched Using the Full Model		
	All Obs. Mean Std. Dev.	Big 4 Mean Std. Dev.	Non-Big 4 Mean Std. Dev.	Difference in Means (t-statistic)	Big 4 Mean Std. Dev.	Non-Big 4 Mean Std. Dev.	Difference in Means (t-statistic)
ADA	0.1031	0.0941	0.1430	−0.0488***	0.1278	0.1312	−0.0034
	0.1398	0.1193	0.2033	(−36.76)	0.1582	0.1918	(−1.42)
Total Assets	1,698	2,021	258	1,763***	333	321	11.77
	5,333	5,789	1,794	(34.71)	2,042	2,006	(0.42)
Total MKT Value	1,872	2,239	234	2,004***	347	289	57.51**
	7,334	8,028	1,694	(28.63)	2,158	1,895	(2.06)
LOG_MKT	4.9784	5.4111	3.0450	2.3660***	3.5247	3.3265	0.1982***
	2.3611	2.2334	1.9013	(110.00)	1.9161	1.8949	(7.56)
ROA	−0.0273	−0.0106	−0.1018	0.0912***	−0.0822	−0.0843	0.0021
	0.2557	0.2206	0.3652	(37.50)	0.2990	0.3411	(0.49)
LEV	0.2159	0.2207	0.1948	0.0261***	0.1994	0.1983	0.0011
	0.1979	0.1967	0.2016	(13.62)	0.2074	0.1988	(0.41)
CURR	3.1482	3.0324	3.6661	−0.6336***	3.5748	3.6350	−0.0602
	5.5247	5.0077	7.3841	(−11.96)	5.5612	7.6348	(−0.66)
No. Obs.	72,600	59,323	13,277		10,565	10,565	
% of Total	100%	77.6%	22.4%				

*, **, *** Indicate significance at the 0.10, 0.05 and 0.01 levels, respectively, using two-tailed t-tests of differences in means.

This table presents the descriptive statistics for our full and propensity-score matched discretionary accruals samples. Propensity scores were calculated using Equation (1).

Variable Definitions:

ADA = absolute discretionary accruals as per Kothari et al. (2005) in year_t;
Total Assets = total assets at the end of year_t;
Total MKT Value = market value of equity at the end of year_t;
LOG_MKT = natural logarithm of market value of equity at the end of year_t;
ROA = income before extraordinary items_{t-1}/average total assets_{t-1};
LEV = (long-term debt_{t-1} plus debt in current liabilities_{t-1})/average total assets_{t-1}; and
CURR = current assets_{t-1}/current liabilities_{t-1}.

propensity-score matched sample, except *LOG_MKT*, are insignificantly different at the 10 percent level between the two client types. *LOG_MKT* is still significantly different between the two client groups because the propensity-score model uses total assets rather than market value to proxy for firm size. When we use market value as the proxy for firm size in the propensity-score model, *LOG_MKT* is insignificantly different between the two client groups.

In Table 2, we confirm the results of Becker et al. (1998), Francis et al. (1999), and Butler et al. (2004) for the full sample, as we find a negative and significant difference in means for *BIG4* of −0.0488 ($t = -36.76$; $p < 0.01$), and a negative and significant *BIG4* coefficient of −0.0179 ($t = -8.48$; $p < 0.01$) in the two Full Sample columns, respectively. All control variable coefficients are significant ($p < 0.01$) and have directional effects consistent with those documented by the previous studies. Regressions in all analyses include untabulated year and industry fixed effects.

TABLE 2

Discretionary Accruals Analysis: Univariate and Multivariate Tests
Full and Propensity-Score Matched Samples

$$ADA_{i,t} = \beta_0 + \beta_1 BIG4_{i,t} + \beta_2 LOG\_MKT_{i,t} + \beta_3 ROA_{i,t-1} + \beta_4 LEV_{i,t-1} + \beta_5 CURR_{i,t-1} + Industry\_FE + Year\_FE + \varepsilon_{i,t} \quad (2)$$

		Dependent Variable: *ADA*			
		Full Sample		Propensity-Score Matched Sample: Matched Using the Full Model	
	Predicted Sign	Difference in Means (t-statistic)	Multivariate Estimate (t-statistic)	Difference in Means (t-statistic)	Multivariate Estimate (t-statistic)
Intercept			0.1128***		0.0828***
			(13.43)		(6.34)
BIG4	−	−0.0488***	−0.0179***	−0.0034	−0.0018
		(−36.76)	(−8.48)	(−1.42)	(−0.73)
LOG_MKT	−		−0.0059***		−0.0054***
			(−18.93)		(−7.99)
ROA	−		−0.1290***		−0.1319***
			(−18.89)		(−15.03)
LEV	−		−0.0497***		−0.0570***
			(−17.08)		(−9.42)
CURR	−		−0.0003***		−0.0006***
			(−2.60)		(−3.43)
Industry_FE			Included		Included
Year_FE			Included		Included
Adjusted R^2			0.14		0.11
No. Obs.		72,600	72,600	21,130	21,130
Matching Model R^2				0.29	
% Correctly Classified				86.03%	

*, **, *** Indicate significance at the 0.10, 0.05 and 0.01 levels, respectively, using two-tailed tests.

This table presents the results of our discretionary accruals tests using the full and propensity-score matched samples. Multivariate estimates are based on Equation (2). Propensity scores were calculated using Equation (1). t-statistics and p-values are calculated using clustered standard errors by firm for the multivariate analyses. For brevity, the year-specific and industry-specific intercepts are not reported. The matching model R^2 is the pseudo R^2 for the propensity-score logistic regression. The percentage correctly classified refers to the percentage of audit clients that are correctly classified as Big 4 or non-Big 4 clients, based on a 50 percent cutoff level, using the predicted probabilities from the propensity-score model.

Variable Definitions:

ADA = absolute discretionary accruals as per Kothari et al. (2005) in $year_t$;
BIG4 = 1 if the client has a Big 4 auditor in $year_t$, and 0 otherwise;
LOG_MKT = natural logarithm of market value of equity at the end of $year_t$;
ROA = income before extraordinary $items_{t-1}$ / average total $assets_{t-1}$;
LEV = (long-term $debt_{t-1}$ plus debt in current $liabilities_{t-1}$)/average total $assets_{t-1}$; and
CURR = current $assets_{t-1}$ / current $liabilites_{t-1}$.

The last two columns of Table 2 present the results of the propensity-score matched sample. We find an insignificant difference in means for *BIG4* of -0.0034 (t = -1.42; p = 0.15) and an insignificant multivariate *BIG4* coefficient of -0.0018 (t = -0.73; p = 0.46), suggesting that once client characteristics are balanced between the two clienteles, the treatment effects of Big 4 auditors are insignificantly different from those of non-Big 4 auditors with respect to discretionary accruals. All control variable coefficients are significant ($p < 0.01$) and have directional effects consistent with those documented in previous studies.[12]

To examine whether the Big 4 effect can be attributed to specific client characteristics, we perform the following analysis. Using the RESET test, we find that total assets (*LOG_ASSETS*), return on assets (ROA_{t-1}), total leverage (LEV_{t-1}), and the current ratio ($CURR_{t-1}$) are nonlinear to both the auditor choice and discretionary accruals; hence, we create single-variable matched samples, with year and industry indicators, using each of these foregoing variables. Table 3 reports the discretionary accruals regressions for these matched samples, where columns one and two report results for the *LOG_ASSETS* matched sample, columns three and four report the results for the ROA_{t-1} matched sample, columns five and six report the results for the LEV_{t-1} matched sample, and columns seven and eight report the results for the $CURR_{t-1}$ matched sample. The *LOG_ASSETS* matched sample has statistically insignificant *BIG4* coefficients of -0.0030 (t = -1.21; p = 0.23) and -0.0026 (t = -1.03; p = 0.30) for the difference in means and multivariate results, respectively, consistent with the results using the full propensity-score matched sample. However, the coefficients on *BIG4* are negative and significant ($p < 0.01$) for all other single-variable matched samples.

It is not surprising that only the *LOG_ASSETS* matched sample yields the same inferences as the full propensity-score matched sample, given that client size is the characteristic most closely related to auditor selection. For example, in Table 2, the full propensity-score matching model explains 29 percent of the variation in auditor choice (see the matching model R^2); whereas in Table 3, the *LOG_ASSETS* matching model (predicting auditor choice with only total assets, year, and industry) explains 28 percent of the variation in auditor choice. The other single-variable matched models at best only explain 6 percent of the variation in auditor choice. Taken together, the findings from our matched samples suggest that the Big 4 effect, using discretionary accruals as an audit-quality proxy, appears to be attributable to client characteristics, primarily client size.

VI. ANALYSIS 2: *EX ANTE* COST-OF-EQUITY CAPITAL

Method

Following Khurana and Raman (2004; hereafter KR), we examine the relation between auditor type and the *ex ante* cost of capital, as follows:

$$RPEG_{i,t} = \beta_0 + \beta_1 BIG4_{i,t} + \beta_2 BETA_{i,t} + \beta_3 LOG\_LEV_{i,t} + \beta_4 VAR_{i,t} + \beta_5 LOG\_MKT_{i,t} + \beta_6 LOG\_BTM_{i,t} + \beta_7 GRWTH_{i,t} + Industry\_FE + Year\_FE + \varepsilon_{i,t}, \quad (3)$$

[12] Hribar and Nichols (2007) document that measurement error in absolute discretionary accruals is often correlated with the partitioning variable in tests of earnings management. They specifically identify that the *BIG4* partitioning variable is highly correlated with the following two main drivers of measurement error in absolute discretionary accruals: firm size and operating volatility (Hribar and Nichols 2007, 1033). Following their guidance to correct for measurement error in estimating the partitioning variable, we simultaneously include several firm-size (the logarithm of assets, sales, and market value) and operating-volatility (cash-flow and sales volatility) controls in our discretionary accruals regressions (untabulated), and find similar results in the full and matched samples to those documented in Tables 2 and 3. Moreover, we ran our analyses separately for positive and negative discretionary accruals and again, all inferences are the same as those documented in our main analyses.

TABLE 3

Discretionary Accruals Analysis: Univariate and Multivariate Tests
Attribute-Based Matched Samples

$$ADA_{i,t} = \beta_0 + \beta_1 BIG4_{i,t} + \beta_2 LOG\_MKT_{i,t} + \beta_3 ROA_{i,t-1} + \beta_4 LEV_{i,t-1} + \beta_5 CURR_{i,t-1} + Industry\_FE + Year\_FE + \varepsilon_{i,t} \quad (2)$$

Dependent Variable: *ADA*

	Size Matched Sample: Matched Using *LOG_ASSETS*, Year, and Industry		ROA_{t-1} Matched Sample: Matched Using ROA_{t-1}, Year, and Industry		LEV_{t-1} Matched Sample: Matched Using LEV_{t-1}, Year, and Industry		$CURR_{t-1}$ Matched Sample: Matched Using $CURR_{t-1}$, Year, and Industry	
	Difference in Means (t-statistic)	**Multivariate Estimate (t-statistic)**	**Difference in Means (t-statistic)**	**Multivariate Estimate (t-statistic)**	**Difference in Means (t-statistic)**	**Multivariate Estimate (t-statistic)**	**Difference in Means (t-statistic)**	**Multivariate Estimate (t-statistic)**
BIG4	−0.0030	−0.0026	−0.0321***	−0.0162***	−0.0436***	−0.0164***	−0.0454***	−0.0174***
	(−1.21)	(−1.03)	(−14.49)	(−5.52)	(−21.16)	(−6.16)	(−22.44)	(−6.99)
Adjusted R^2		0.11		0.13		0.13		0.13
No. Obs.	22,210	22,210	26,456	26,456	26,540	26,540	26,552	26,552
Matching Model R^2	0.28		0.06		0.04		0.04	
% Correctly Classified	85.78%		81.89%		81.72%		81.71%	

*, **, *** Indicate significance at the 0.10, 0.05 and 0.01 levels, respectively, using two-tailed tests.

This table presents the results of our discretionary accruals tests using four different attribute-based matched samples. We use the propensity-score methodology to implement the attribute-based matching. Propensity scores were calculated using only a single variable with year and industry indicators. Multivariate estimates are based on Equation (2). t-statistics and p-values are calculated using clustered standard errors by firm for the multivariate analyses. For brevity, we only report the coefficients on *BIG4*. The matching model R^2 is the pseudo R^2 for the propensity-score logistic regression. The percentage correctly classified refers to the percentage of audit clients that are correctly classified as Big 4 or non-Big 4 clients, based on a 50 percent cutoff level, using the predicted probabilities from the propensity-score model.

Variable Definitions:

ADA = absolute discretionary accruals as per Kothari et al. (2005) in $year_t$;
BIG4 = 1 if the client has a Big 4 auditor in $year_t$, and 0 otherwise;
LOG_MKT = natural logarithm of market value of equity at the end of $year_t$;
ROA = income before extraordinary $items_{t-1}$ / average total $assets_{t-1}$;
LEV = (long-term $debt_{t-1}$ plus debt in current $liabilities_{t-1}$)/average total $assets_{t-1}$; and
CURR = current $assets_{t-1}$ / current $libalities_{t-1}$.

where for firm *i* and fiscal year *t*:

RPEG = *ex ante* cost-of-equity capital estimated using the r_{peg} approach as per Easton (2004) and used by KR;
BIG4 = 1 if the client has a Big 4 auditor in $year_t$, and 0 otherwise;
BETA = stock beta (systematic risk) calculated over the 36 months ending in the month of the fiscal $year\text{-}end_t$;
LOG_LEV = natural logarithm of (long-term $debt_t$ plus debt in current $liabilities_t$)/average total $assets_{t-1}$;
VAR = earnings variability measured by the dispersion in analysts' earnings forecasts available on I/B/E/S during the fiscal year-end month;
LOG_MKT = natural logarithm of market value of equity at the end of $year_t$;
LOG_BTM = natural logarithm of book value of equity/market value of equity at the end of $year_t$; and
GRWTH = forecasted growth measured as the difference between the mean analysts' two-year- and one-year-ahead earnings forecasts scaled by the one-year-ahead earnings forecast.

For comparative purposes, all variables and their predicted directions are as specified by KR.

Results

Table 4 presents descriptive statistics for both the full and propensity-score matched *ex ante* cost-of-equity capital, *RPEG*, samples. Due to restrictions concerning analyst forecast data, there are 25,068 firm-year observations in the full sample, of which 23,856 (94.9 percent) and 1,212 (5.1 percent) represent Big 4 and non-Big 4 clients, respectively. On average, firms in this full sample are significantly larger across both auditor types than in the accruals full sample: the mean market value and total assets of Big 4 clients are \$4.5 billion and \$4.2 billion, respectively, and the mean market value and total assets of non-Big 4 clients are \$1.5 billion and \$1.9 billion, respectively, indicating that the size differential is smaller in this sample than in the accruals sample. Also worth noting in the *RPEG* full sample is that Big 4 clients, on average, have significantly lower *RPEG*, *BETA*, and *LOG_BTM* values than non-Big 4 clients. The propensity-score model once again appears effective in forming a balanced sample of Big 4 and non-Big 4 clients as all the *RPEG* control variables in the propensity-score matched sample are insignificantly different at the 10 percent level between the two client types.

In Table 5, we first confirm the results of KR for the full sample, as we find a negative and significant difference in means for *BIG4* of -0.0072 ($t = -4.99$; $p < 0.01$), and a negative and significant *BIG4* coefficient of -0.0037 ($t = -2.14$, $p = 0.03$) in the two Full Sample columns, respectively. All control variable coefficients are significant ($p < 0.01$) and have directional effects consistent with those documented by KR for their U.S. sample. The last two columns of Table 5 present the results of the propensity-score matched sample. We find an insignificant difference in means for *BIG4* of -0.0031 ($t = -1.43$; $p = 0.15$) and an insignificant *BIG4* coefficient of -0.0021 ($t = -1.01$; $p = 0.31$). All the control variable coefficients remain significant and have directional effects consistent with those reported by KR.

Again, to investigate whether the Big 4 effect could reflect client characteristics, we perform the following procedures. Using the RESET test, we find that total assets (*LOG_ASSETS*), the stock beta (*BETA*), total leverage (*LOG_LEV*), and the book-to-market ratio (*LOG_BTM*) are nonlinear to both the auditor choice and to the *ex ante* cost-of-equity capital. Hence, we create single-variable matched samples, with year and industry indicators, using each of these variables. Table 6 reports the *RPEG* regressions using these attribute-based matched samples, where col-

TABLE 4

Ex Ante Cost-of-Equity Capital Analysis: Descriptive Statistics Full and Propensity-Score Matched Samples

	Full Sample				Propensity-Score Matched Sample: Matched Using the Full Model		
	All Obs. Mean Std. Dev.	Big 4 Mean Std. Dev.	Non-Big 4 Mean Std. Dev.	Difference in Means (t-statistic)	Big 4 Mean Std. Dev.	Non-Big 4 Mean Std. Dev.	Difference in Means (t-statistic)
RPEG	0.1146	0.1142	0.1215	−0.0072***	0.1182	0.1214	−0.0031
	0.0492	0.0489	0.0537	(−4.99)	0.0530	0.0536	(−1.43)
Total Assets	4,093	4,202	1,944	2,258***	1,723	1,955	−231.51
	12,122	12,258	8,784	(6.33)	7,802	8,812	(−0.68)
Total MKT Value	4,352	4,499	1,466	3,033***	1,516	1,474	42.17
	12,342	12,575	5,428	(8.36)	5,906	5,444	(0.18)
LOG_MKT	6.7469	6.7938	5.8235	0.9702***	5.8871	5.8257	0.0614
	1.7851	1.7837	1.5471	(18.59)	1.5574	1.5518	(0.97)
BETA	1.1029	1.0959	1.2403	−0.1443***	1.1983	1.2441	−0.0461
	0.7566	0.7461	0.9292	(−6.48)	0.8072	0.9308	(−1.28)
LOG_LEV	−1.8453	−1.8283	−2.1791	0.3510***	−2.2303	−2.1860	−0.0442
	1.3257	1.3060	1.6310	(9.01)	1.7432	1.6341	(−0.64)
VAR	0.0687	0.0689	0.0658	0.0030	0.0640	0.0658	−0.0018
	0.1033	0.1037	0.0961	(1.01)	0.0101	0.0960	(−0.45)
LOG_BTM	−0.8294	−0.8331	−0.7565	−0.0766***	−0.7405	−0.7591	0.0186
	0.6862	0.6883	0.6374	(−3.79)	0.6928	0.6383	(0.69)
GRWTH	0.3314	0.3306	0.3476	−0.0171	0.4150	0.3479	0.0671
	1.5853	1.6051	1.1257	(−0.37)	2.0604	1.1289	(0.99)
No. Obs.	25,068	23,856	1,212		1,204	1,204	
% of Total	100%	94.9%	5.1%				

*, **, *** Indicate significance at the 0.10, 0.05 and 0.01 levels, respectively, using two-tailed t-tests of differences in means.

This table presents the descriptive statistics for our full and propensity-score matched *ex ante* cost-of-equity capital samples. Propensity scores were calculated using Equation (1).

Variable Definitions:

RPEG = *ex ante* cost-of-equity capital estimated using the r_{peg} approach as per Easton (2004) and used by Khurana and Raman (2004);
Total Assets = total assets at the end of $year_t$;
Total MKT Value = market value of equity at the end of $year_t$;
LOG_MKT = natural logarithm of market value of equity at the end of $year_t$;
BETA = stock beta (systematic risk) calculated over the 36 months ending in the month of the fiscal $year\text{-}end_t$;
LOG_LEV = natural logarithm of (long-term $debt_t$ plus debt in current $liabilities_t$)/average total $assets_{t-1}$;
VAR = earnings variability measured by the dispersion in analysts' earnings forecasts available on I/B/E/S during the fiscal year-end month;
LOG_BTM = natural logarithm of book value of equity/market value of equity at the end of $year_t$; and
GRWTH = forecasted growth measured as the difference between the mean analysts' two-year- and one-year-ahead earnings forecasts scaled by the one-year-ahead earnings forecast.

TABLE 5

Ex Ante Cost-of-Equity Capital Analysis: Univariate and Multivariate Tests Full and Propensity-Score Matched Samples

$$RPEG_{i,t} = \beta_0 + \beta_1 BIG4_{i,t} + \beta_2 BETA_{i,t} + \beta_3 LOG\_LEV_{i,t} + \beta_4 VAR_{i,t} + \beta_5 LOG\_MKT_{i,t} + \beta_6 LOG\_BTM_{i,t} + \beta_7 GRWTH_{i,t} + Industry\_FE + Year\_FE + \varepsilon_{i,t} \quad (3)$$

		Dependent Variable: *RPEG*			
		Full Sample		Propensity-Score Matched Sample: Matched Using the Full Model	
	Predicted Sign	Difference in Means (t-statistic)	Multivariate Estimate (t-statistic)	Difference in Means (t-statistic)	Multivariate Estimate (t-statistic)
Intercept			0.1558***		0.1517***
			(22.09)		(6.15)
BIG4	−	−0.0072***	−0.0037**	−0.0031	−0.0021
		(−4.99)	(−2.14)	(−1.43)	(−1.01)
BETA	+		0.0099***		0.0091***
			(16.51)		(5.29)
LOG_LEV	+		0.0029***		0.0028***
			(10.52)		(4.40)
VAR	+		0.0232***		0.0313**
			(5.38)		(2.14)
LOG_MKT	−		−0.0062***		−0.0093***
			(−19.77)		(−8.96)
LOG_BTM	+		0.0105***		0.0048**
			(13.71)		(1.99)
GRWTH	+		0.0067***		0.0079***
			(4.63)		(5.04)
Industry_FE			Included		Included
Year_FE			Included		Included
Adjusted R^2			0.28		0.33
No. Obs.		25,068	25,068	2,408	2,408
Matching Model R^2				0.11	
% Correctly Classified				95.00%	

*, **, *** Indicate significance at the 0.10, 0.05 and 0.01 levels, respectively, using two-tailed tests.

This table presents the results of our *ex ante* cost-of-equity capital tests using the full and propensity-score matched samples. Multivariate estimates are based on Equation (3). Propensity scores were calculated using Equation (1). t-statistics and p-values are calculated using clustered standard errors by firm for the multivariate analyses. For brevity, the year-specific and industry-specific intercepts are not reported. The matching model R^2 is the pseudo R^2 for the propensity-score logistic regression. The percentage correctly classified refers to the percentage of audit clients that are correctly classified as Big 4 or non-Big 4 clients, based on a 50 percent cutoff level, using the predicted probabilities from the propensity-score model.

Variable Definition:

RPEG = *ex ante* cost-of-equity capital estimated using the r_{peg} approach as per Easton (2004) and used by Khurana and Raman (2004);

BIG4 = 1 if the client has a Big 4 auditor in $year_t$, and 0 otherwise;

(continued on next page)

TABLE 5 (continued)

BETA = stock beta (systematic risk) calculated over the 36 months ending in the month of the fiscal year-end$_t$;
LOG_LEV = natural logarithm of (long-term debt$_t$ plus debt in current liabilities$_t$)/average total assets$_{t-1}$;
VAR = earnings variability measured by the dispersion in analysts' earnings forecasts available on I/B/E/S during the fiscal year-end month;
LOG_MKT = natural logarithm of market value of equity at the end of year$_t$;
LOG_BTM = natural logarithm of book value of equity/market value of equity at the end of year$_t$; and
GRWTH = forecasted growth measured as the difference between the mean analysts' two-year- and one-year-ahead earnings forecasts scaled by the one-year-ahead earnings forecast.

umns one and two report results for the *LOG_ASSETS* matched sample, columns three and four report the results for the *BETA* matched sample, columns five and six report the results for the *LOG_LEV* matched sample, and columns seven and eight report the results for the *LOG_BTM* matched sample.

As in the accruals analysis, the difference in means and the coefficient on *BIG4* for the multivariate results in the *LOG_ASSETS* matched sample are insignificant: −0.0032 ($t = -1.49$; $p = 0.14$) and −0.0013 ($t = -0.61$; $p > 0.50$), respectively. The differences in means for *BIG4* are negative and significant, and the coefficients on *BIG4* in the multivariate results are insignificant for all other single-variable matched samples. As in the discretionary accruals analysis, the *LOG_ASSETS* matched sample yields the same inferences as the full propensity-score matched sample. Together, these findings suggest that the Big 4 effect on the *ex ante* cost-of-equity capital are likely attributable to client characteristics, primarily client size.

VII. ANALYSIS 3: ANALYST FORECAST ACCURACY

Method

To test whether the Big 4 versus non-Big 4 differences in analyst forecast accuracy can be attributed to client characteristics, we use the following model employed by Behn et al. (2008; hereafter BCK):

$$\begin{aligned} ACCY_{i,t} = {} & \beta_0 + \beta_1 BIG4_{i,t} + \beta_2 LOG\_MKT_{i,t} + \beta_3 SURPRISE_{i,t} + \beta_4 NETLOSS_{i,t} + \beta_5 ZMIJ_{i,t} \\ & + \beta_6 HORIZON_{i,t} + \beta_7 STDROE_{i,t} + \beta_8 NANA_{i,t} + \beta_9 EL_{i,t} + Industry\_FE + Year\_FE \\ & + \varepsilon_{i,t}, \end{aligned} \tag{4}$$

where for firm *i* and fiscal year *t*:

ACCY = negative of the absolute value of the difference between analysts' earnings forecasts of eps$_{t+1}$ and actual eps$_{t+1}$, scaled by stock price at the end of year$_t$, as per Lang and Lundholm (1996) and used by BCK;
BIG4 = 1 if the client has a Big 4 auditor in year$_t$, and 0 otherwise;
LOG_MKT = natural logarithm of market value of equity at the end of year$_{t+1}$;
SURPRISE = (net income$_{t+1}$ − net income$_t$)/market value of equity at the end of year$_t$;
NETLOSS = 1 if client has negative net income$_{t+1}$, and 0 otherwise;
ZMIJ = distress score calculated using Zmijewski's (1984) unweighted original parameters;
HORIZON = natural logarithm of the average number of calendar days between forecast announcement date and subsequent earnings announcement date;
STDROE = standard deviation of net income over the five years from year$_{t-4}$ until year$_t$;
NANA = natural logarithm of the number of analysts following the client; and

TABLE 6

Ex Ante Cost-of-Equity Capital Analysis: Univariate and Multivariate Tests Attribute-Based Matched Samples

$$RPEG_{i,t} = \beta_0 + \beta_1 BIG4_{i,t} + \beta_2 BETA_{i,t} + \beta_3 LOG\_LEV_{i,t} + \beta_4 VAR_{i,t} + \beta_5 LOG\_MKT_{i,t} + \beta_6 LOG\_BTM_{i,t} + \beta_7 GRWTH_{i,t} + Industry\_FE + Year\_FE + \varepsilon_{i,t} \quad (3)$$

Dependent Variable: _RPEG_

	Size Matched Sample: Matched Using _LOG_ASSETS_, Year and Industry		_BETA_ Matched Sample: Matched Using _BETA_, Year and Industry		_LOG_LEV_ Matched Sample: Matched Using _LOG_LEV_, Year and Industry		BTM Ratio Matched Sample: Matched Using BTM Ratio, Year and Industry	
	Difference in Means (t-statistic)	Multivariate Estimate (t-statistic)	Difference in Means (t-statistic)	Multivariate Estimate (t-statistic)	Difference in Means (t-statistic)	Multivariate Estimate (t-statistic)	Difference in Means (t-statistic)	Multivariate Estimate (t-statistic)
BIG4	−0.0032	−0.0013	−0.0087***	−0.0011	−0.0119***	−0.0025	−0.0106***	−0.0015
	(−1.49)	(−0.61)	(−4.18)	(−0.54)	(−5.96)	(−1.32)	(−5.17)	(−0.72)
Adjusted R^2		0.33		0.33		0.32		0.34
No. Obs.	2,408	2,408	2,408	2,408	2,408	2,408	2,408	2,408
Matching Model R^2	0.11		0.05		0.06		0.05	
% Correctly Classified	95.02%		95.01%		95.01%		95.01%	

*, **, *** Indicate significance at the 0.10, 0.05 and 0.01 levels, respectively, using two-tailed tests.

This table presents the results of our *ex ante* cost-of-equity capital tests using four different attribute-based matched samples. We use the propensity-score methodology to implement the attribute-based matching. Propensity scores were calculated using only a single variable with year and industry indicators.

Multivariate estimates are based on Equation (3). t-statistics and p-values are calculated using clustered standard errors by firm for the multivariate analyses.

For brevity, we only report the coefficients on *BIG4*. The matching model R^2 is the pseudo R^2 for the propensity-score logistic regression. The percentage correctly classified refers to the percentage of audit clients that are correctly classified as Big 4 or non-Big 4 clients, based on a 50 percent cutoff level, using the predicted probabilities from the propensity-score model.

Variable Definitions:

RPEG = *ex ante* cost-of-equity capital estimated using the r_{peg} approach as per Easton (2004) and used by Khurana and Raman (2004);

BIG4 = 1 if the client has a Big 4 auditor in $year_t$, and 0 otherwise;

(continued on next page)

TABLE 6 (continued)

BETA = stock beta (systematic risk) calculated over the 36 months ending in the month of the fiscal year-end$_t$;
LOG_LEV = natural logarithm of (long-term debt$_t$ plus debt in current liabilities$_t$)/average total assets$_{t-1}$;
VAR = earnings variability measured by the dispersion in analysts' earnings forecasts available on I/B/E/S during the fiscal year-end month;
LOG_MKT = natural logarithm of market value of equity at the end of year$_t$;
LOG_BTM = natural logarithm of book value of equity/market value of equity at the end of year$_t$; and
GRWTH = forecasted growth measured as the difference between the mean analysts' two-year- and one-year-ahead earnings forecasts scaled by the one-year-ahead earnings forecast.

EL = actual eps$_{t+1}$.

For comparative purposes, all variables and their predicted directions are as specified by BCK.

Results

Table 7 presents descriptive statistics for both the full and propensity-score matched analyst forecast accuracy, *ACCY*, samples. The full sample has a total of 28,037 firm-year observations, of which 26,521 (94.3 percent) and 1,516 (5.7 percent) represent Big 4 and non-Big 4 clients, respectively. As with the *ex ante* cost-of-equity capital analysis, the firms in this full sample are significantly larger across both client types than in the accruals full sample and, hence, the size differential between the two groups is again smaller in this sample than it is in the accruals analysis. Overall, the descriptive statistics again confirm that Big 4 and non-Big 4 clienteles are significantly different in terms of size and profitability. Moreover, relevant to this analysis, we find that Big 4 audit clients have a larger analyst following and a lower probability of bankruptcy than non-Big 4 audit clients.

In Table 8, we first confirm the results of BCK for the full sample, as we find a positive and significant difference in means for *BIG4* of 0.0083 ($t = 7.21$; $p < 0.01$), and a positive and significant *BIG4* coefficient of 0.0042 ($t = 2.34$; $p = 0.02$) in the two Full Sample columns, respectively. Except for *HORIZON*, *NANA*, and *EL*, all control variable coefficients have directional effects and significance levels consistent with those documented by BCK. The last two columns of Table 8 present the results of the propensity-score matched sample. We find an insignificant difference in means for *BIG4* of 0.0026 ($t = 1.30$; $p = 0.19$) and an insignificant *BIG4* coefficient of 0.0031 ($t = 1.54$; $p = 0.12$).

Using the RESET test, we find that the total assets *(LOG_ASSETS)*, the standard deviation of the return on equity (*STDROE*), and analyst following (*NANA*) are nonlinear to both the auditor choice and to analyst forecast accuracy. Hence, we create single-variable matched samples, with year and industry indicators, using each of these variables. Table 9 reports the *ACCY* regressions for the attribute-based matched samples, where columns one and two report results for the *LOG_ASSETS* matched sample, columns three and four report the results for the *STDROE* matched sample, and columns five and six report the results for the *NANA* matched sample.

As in the accruals and *ex ante* cost-of-equity capital analyses, the difference in means and the coefficients on *BIG4* for the multivariate results in the *LOG_ASSETS* matched sample are insignificant: -0.0021 ($t = -0.91$; $p = 0.36$) and 0.0013 ($t = 0.59$; $p > 0.50$), respectively. For the *STDROE* matched sample, the difference in means for *BIG4* is positive and significant, while the *BIG4* coefficient in the multivariate analyses is insignificant. Moreover, the difference in means and the coefficients on *BIG4* in the *NANA* matched sample are positive and significant. In line with the other two analyses, these findings suggest that client characteristics, and particularly client size, could potentially confound the inferences pertaining to the Big 4 effect.

Our findings concerning analyst forecast accuracy appear to be consistent with a 2008 CFA Institute Center survey of 617 CFA investment analysts documenting that the majority of analysts do not prefer Big 4 auditors to non-Big 4 auditors. Specifically, only 41 percent of the respondents generally indicated that they had a preference for firms using "brand-name" auditors; moreover, only 15 percent of the respondents thought that the attractiveness of a company as an investment is detracted when a smaller company switches to a lower-cost auditor that may be more efficient and cost-effective (CFA Institute Center 2008).

VIII. SENSITIVITY ANALYSES

Bootstrapping, Kernel Weighting, and Random Subsamples

To mitigate concerns that our findings are a consequence of the smaller sample sizes, we

TABLE 7

Analyst Forecast Accuracy Analysis: Descriptive Statistics
Full and Propensity-Score Matched Samples

	Full Sample				Propensity-Score Matched Sample: Matched Using the Full Model		
	All Obs. Mean Std. Dev.	Big 4 Mean Std. Dev.	Non-Big 4 Mean Std. Dev.	Difference in Means (t-statistic)	Big 4 Mean Std. Dev.	Non-Big 4 Mean Std. Dev.	Difference in Means (t-statistic)
ACCY	−0.0122	−0.0117	−0.0201	0.0083***	−0.0162	−0.0189	0.0027
	0.0437	0.0425	0.0608	(7.21)	0.0511	0.0593	(1.30)
Total Assets	4,243	4,377	1,893	2,484***	1,382	1,929	−547**
	12,787	12,990	8,141	(7.36)	5,527	8,245	(−2.12)
Total MKT Value	4,875	5,058	1,677	3,381***	1,238	1,708	−470**
	17,331	17,729	6,756	(7.39)	4,367	6,839	(−2.22)
LOG_MKT	6.6934	6.7582	5.5594	1.1989***	5.6056	5.5848	0.0208
	1.8677	1.8508	1.7975	(24.57)	1.6483	1.7950	(0.33)
SURPRISE	0.0814	0.0812	0.0838	−0.0026	0.0865	0.0821	0.0045
	0.1917	0.1909	0.2062	(−0.50)	0.2198	0.2040	(0.58)
NETLOSS	0.2036	0.2029	0.2170	−0.0141*	0.2203	0.2108	0.0095
	0.4027	0.4021	0.4123	(−1.33)	0.4146	0.4080	(0.63)
ZMIJ	−3.0792	−3.0710	−3.2154	0.1441***	−3.1456	−3.2279	0.0822
	1.3155	1.3085	1.4266	(4.15)	1.4487	1.4024	(1.57)
HORIZON	3.3829	3.3833	3.3765	0.0068	3.3892	3.3766	0.0126
	0.5051	0.5029	0.5426	(0.51)	0.5385	0.5372	(0.74)
STDROE	0.3623	0.3486	0.6025	−0.2539***	0.5285	0.5580	−0.0295
	1.1282	1.0816	1.7360	(−8.53)	1.4723	1.6361	(−0.52)
NANA	0.9265	0.9457	0.5905	0.3552***	0.5940	0.5985	−0.0045
	0.8588	0.8626	0.7106	(15.73)	0.6943	0.7137	(−0.17)
EL	0.9682	0.9798	0.7661	0.2137***	0.7489	0.7862	−0.0372
	2.0088	2.0342	1.4804	(4.03)	1.5335	1.4806	(−0.66)
No. Obs.	28,037	26,521	1,516		1,475	1,475	
% of Total	100%	94.3%	5.7%				

*, **, *** Indicate significance at the 0.10, 0.05 and 0.01 levels, respectively, using two-tailed t-tests of differences in means.

This table presents the descriptive statistics for our full and propensity-score matched analyst forecast-accuracy samples. Propensity scores were calculated using Equation (1).

Variable Definitions:

ACCY = negative of the absolute value of the difference between analysts' earnings forecasts of eps_{t+1} and actual eps_{t+1}, scaled by stock price at the end of $year_t$, as per Lang and Lundholm (1996) and used by Behn et al. (2008);
Total Assets = total assets at the end of $year_{t+1}$;
Total MKT Value = market value of equity at the end of $year_{t+1}$;
LOG_MKT = natural logarithm of market value of equity at the end of $year_{t+1}$;
SURPRISE = ($net\ income_{t+1}$ − $net\ income_t$)/market value of equity at the end of $year_t$;
NETLOSS = 1 if the client has negative $net\ income_{t+1}$, and 0 otherwise;
ZMIJ = distress score calculated using Zmijewski's (1984) unweighted original parameters;
HORIZON = natural logarithm of the average number of calendar days between the forecast announcement date and the subsequent earnings announcement date;
STDROE = standard deviation of net income over the five years from $year_{t-4}$ until $year_t$;
EL = actual eps_{t+1}; and
NANA = natural logarithm of the number of analysts following the client.

TABLE 8

Analyst Forecast Accuracy Analysis: Univariate and Multivariate Tests Full and Propensity-Score Matched Samples

$$ACCY_{i,t} = \beta_0 + \beta_1 BIG4_{i,t} + \beta_2 LOG\_MKT_{i,t} + \beta_3 SURPRISE_{i,t} + \beta_4 NETLOSS_{i,t} + \beta_5 ZMIJ_{i,t} + \beta_6 \mathrm{HORIZON}_{i,t} + \beta_7 \mathrm{STDROE}_{i,t} + \beta_8 NANA_{i,t} + \beta_9 EL_{i,t} + Industry\_FE + Year\_FE + \varepsilon_{i,t} \quad (4)$$

		Dependent Variable: *ACCY*			
		Full Sample		Propensity-Score Matched Sample: Matched Using the Full Model	
	Predicted Sign	Difference in Means (t-statistic)	Multivariate Estimate (t-statistic)	Difference in Means (t-statistic)	Multivariate Estimate (t-statistic)
Intercept			−0.0414*** (−9.08)		−0.0302*** (−2.59)
BIG4	+	0.0083*** (7.21)	0.0042** (2.34)	0.0026 (1.30)	0.0031 (1.54)
LOG_MKT	+		0.0027*** (10.12)		0.0052*** (4.36)
SURPRISE	−		−0.0598*** (−13.43)		−0.0439*** (−4.46)
NETLOSS	−		−0.0072*** (−6.87)		−0.0198*** (−4.68)
ZMIJ	−		−0.0028*** (−6.69)		−0.0024* (−1.79)
HORIZON	−		0.0001 (0.22)		−0.0003 (−0.17)
STDROE	−		−0.0021*** (−4.07)		−0.0035*** (−2.66)
NANA	+		0.0016*** (4.25)		−0.0006 (−0.27)
EL	?		0.0006** (2.38)		−0.0006 (−0.62)
Industry_FE			Included		Included
Year_FE			Included		Included
Adjusted R^2			0.19		0.18
No. Obs.		28,037	28,037	2,950	2,950
Matching Model R^2				0.14	
% Correctly Classified				94.62%	

*, **, *** Indicate significance at the 0.10, 0.05 and 0.01 levels, respectively, using two-tailed tests.

This table presents the results of our analyst forecast accuracy tests using the full and propensity-score matched samples. Multivariate estimates are based on Equation (4). Propensity scores were calculated using Equation (1).

Following Behn et al. (2008), t-statistics and p-values are calculated using White's (1980) heteroscedasticity-corrected standard errors clustered by firm for the multivariate analyses. For brevity, the year-specific and industry-specific intercepts

(continued on next page)

TABLE 8 (continued)

are not reported. The matching model R^2 is the pseudo R^2 for the propensity-score logistic regression. The percentage correctly classified refers to the percentage of audit clients that are correctly classified as Big 4 or non-Big 4 clients, based on a 50 percent cutoff level, using the predicted probabilities from the propensity-score model.

Variable Definitions:

ACCY = negative of the absolute value of the difference between analysts' earnings forecasts of eps_{t+1} and actual eps_{t+1}, scaled by stock price at the end of $year_t$, as per Lang and Lundholm (1996) and used by Behn et al. (2008);
BIG4 = 1 if the client has a Big 4 auditor in $year_t$, and 0 otherwise;
LOG_MKT = natural logarithm of market value of equity at the end of $year_{t+1}$;
SURPRISE = (net $income_{t+1}$ − net $income_t$)/market value of equity at the end of $year_t$;
NETLOSS = 1 if the client has negative net $income_{t+1}$, and 0 otherwise;
ZMIJ = distress score calculated using Zmijewski's (1984) unweighted original parameters;
HORIZON = natural logarithm of the average number of calendar days between the forecast announcement date and the subsequent earnings announcement date;
STDROE = standard deviation of net income over the five years from $year_{t-4}$ until $year_t$;
EL = actual eps_{t+1}; and
NANA = natural logarithm of the number of analysts following the client.

perform the following analyses. First, for each matched sample in our main analyses, we obtain bootstrap estimates, using 10,000 repetitions. Second, using each audit-quality proxy, we run analyses for the full sample, putting more weight and using kernel weighting (Heckman et al. 1997), on the observations that are closer in the propensity score or client size. We find that all inferences from these tests are the same as those documented in our matched sample analyses.

Moreover, to ensure that the Big 4 effect can still be detected in smaller subsamples, we perform the following analyses. First, for each audit-quality proxy, we obtain bootstrap estimates using 10,000 repetitions and sample sizes equal to those of the matched sample in each repetition. Second, for each audit-quality proxy, we initially select a random subsample with the same number of observations and composition (i.e., 50 percent Big 4 clients and 50 percent Non-Big 4 clients) as in the matched sample. Using this random subsample, we obtain bootstrap estimates using 10,000 repetitions.[13] We find that the Big 4 effect holds in these alternative analyses.

Varying the Proportion of Big 4 and Non-Big 4 Clients in the Full and Matched Samples

Some of our audit-quality proxies result in samples that have different auditor compositions than the full population, potentially biasing our inferences. Hence, for each audit-quality proxy, we select a random sample with an 80/20 ratio between Big and non-Big 4 auditors, as in the full Compustat population, from both our full and matched samples, and obtain bootstrap estimates using 10,000 repetitions.[14] All reported inferences are robust to these alternative specifications.

Additional Client and Auditor Characteristics

Facing a similar problem of separating the effects of client from those of audit firm characteristics, Francis and Yu (2009; hereafter FY) document that the audit quality of larger Big 4

[13] For example, the discretionary accruals propensity-score matched sample has 10,565 clients of each auditor type; hence, we randomly select 10,565 of each auditor type from our full sample and estimate our regression Equation (2) for each random sample, repeating this procedure 10,000 times to obtain bootstrapped estimates.

[14] For example, from the discretionary accruals full sample, we randomly draw 40,000 observations from Big 4 clients and 10,000 observations from non-Big 4 clients, achieving an 80/20 ratio. Similarly, from the discretionary accruals matched sample, we randomly draw 8,000 observations from Big 4 clients and 2,000 observations from non-Big 4 clients achieving an 80/20 ratio. Using these two subsamples, one derived from the full sample and the other from the matched sample, we re-estimate our regression models with 10,000 bootstrap repetitions.

TABLE 9

Analyst Forecast Accuracy Analysis: Univariate and Multivariate Tests Attribute-Based Matched Samples

$$ACCY_{i,t} = \beta_0 + \beta_1 BIG4_{i,t} + \beta_2 LOG\_MKT_{i,t} + \beta_3 SURPRISE_{i,t} + \beta_4 NETLOSS_{i,t} + \beta_5 ZMIJ_{i,t} + \beta_6 HORIZON_{i,t} + \beta_7 STDROE_{i,t} + \beta_8 NANA_{i,t} + \beta_9 EL_{i,t} + Industry\_FE + Year\_FE + \varepsilon_{i,t} \quad (4)$$

Dependent Variable: *ACCY*

	Size Matched Sample: Matched Using *LOG_ASSETS*, Year and Industry		*STDROE* Matched Sample: Matched Using *STDROE*, Year and Industry		*NANA* Matched Sample: Matched Using NANA, Year and Industry	
	Difference in Means (t-statistic)	Multivariate Estimate (t-statistic)	Difference in Means (t-statistic)	Multivariate Estimate (t-statistic)	Difference in Means (t-statistic)	Multivariate Estimate (t-statistic)
BIG4	−0.0021	0.0013	0.0078***	0.0024	0.0105***	0.0038***
	(−0.91)	(0.59)	(4.08)	(1.19)	(6.13)	(2.10)
Adjusted R^2		0.21		0.22		0.20
No. Obs.	2,956	2,956	2,982	2,982	2,960	2,960
Matching Model R^2	0.13		0.05		0.06	
% Correctly Classified	94.60%		94.55%		94.56%	

*, **, *** Indicate significance at the 0.10, 0.05 and 0.01 levels, respectively, using two-tailed tests.

This table presents the results of our analyst forecast accuracy tests using three different attribute-based matched samples. We use the propensity-score methodology to implement the attribute-based matching. Propensity scores were calculated using only a single variable with year and industry indicators.

Multivariate estimates are based on Equation (4). Following Behn et al. (2008), t-statistics and p-values are calculated using White's (1980) heteroscedasticity-corrected standard errors clustered by firm for the multivariate analyses. For brevity, we only report the coefficients on *BIG4*. The matching model R^2 is the pseudo R^2 for the propensity-score logistic regression. The percentage correctly classified refers to the percentage of audit clients that are correctly classified as Big 4 or non-Big 4 clients, based on a 50 percent cutoff level, using the predicted probabilities from the propensity-score model.

Variable Definitions:

ACCY = negative of the absolute value of the difference between analysts' earnings forecasts of eps_{t+1} and actual eps_{t+1}, scaled by stock price at the end of $year_t$, as per Lang and Lundholm (1996) and used by Behn et al. (2008);
BIG4 = 1 if the client has a Big 4 auditor in $year_t$, and 0 otherwise;
LOG_MKT = natural logarithm of market value of equity at the end of $year_{t+1}$;
SURPRISE = (net $income_{t+1}$ − net $income_t$)/market value of equity at the end of $year_t$;
NETLOSS = 1 if the client has negative net $income_{t+1}$, and 0 otherwise;
ZMIJ = distress score calculated using Zmijewski's (1984) unweighted original parameters;
HORIZON = natural logarithm of the average number of calendar days between the forecast announcement date and the subsequent earnings announcement date;
STDROE = standard deviation of net income over the five years from $year_{t-4}$ until $year_t$;
EL = actual eps_{t+1}; and
NANA = natural logarithm of the number of analysts following the client.

offices is superior to that of smaller Big 4 offices. We enhance our study's controls by estimating all our propensity-score and audit-quality regressions including FY's comprehensive list of control variables.[15] We find that the coefficients on *BIG4* in the full and matched samples are insignificant for all three audit-quality proxies. These results (untabulated) provide additional support for the argument that the Big 4 distinction may reflect client and not auditor characteristics.[16]

Auditor Office Size and Client Characteristics

FY's tests comparing audit quality among large and small auditor offices may encounter similar difficulties as the Big 4 auditor tests, given that there is a strong positive correlation between office size and client size. To investigate this possibility, we first replicate and confirm FY's study using their full absolute discretionary accruals, small profit increases, and going-concern opinions models, and their sample years. Second, we create an indicator variable, *MEDOFFICE*, equal to 1 for offices above the median office size, and 0 otherwise, in order to match clients of large offices to those of small offices. Third, we confirm FY's findings by replacing *LOGOFFICE* with *MEDOFFICE* in the full samples for the three audit-quality proxies. Finally, we use both client-size and propensity-score matching to match clients of offices above the median office size to those clients of offices below the median office size. We then find an insignificant difference in *MEDOFFICE* between large and small offices for three foregoing audit-quality proxies ($p > 0.50$, $p = 0.29$, and $p = 0.20$, respectively). These matched-sample results are robust to 10,000 bootstrap repetitions. We caution that this analysis does not rule out FY's conclusion that large offices have higher audit quality, although it suggests the importance of fully controlling for client characteristics in tests of audit quality.

Nonlinear Robustness Checks

To some degree, matching models purge the impact of nonlinearities on auditor size in our analyses. Nonetheless, as a robustness check, we use a backfitting partial linear model (Hastie and Tibshirani 1990) to estimate both the linear and nonlinear terms previously identified using the RESET test.[17] We find that all inferences pertaining to the full and matched samples using partial linear regressions are the same as those documented in the main analyses using ordinary least-squares (OLS) regressions.

---

[15] These variables are as follows: *LOGOFFICE* is the natural logarithm of practice office size based on the aggregated client audit fees (in $ millions) of a practice office in a specific fiscal year; *INFLUENCE* is the ratio of a specific client's total fees (audit fees plus nonaudit fees) relative to aggregate annual fees generated by the practice office that audits the client; *TENURE* is 1 if auditor tenure is three years or less, and 0 otherwise; *NATIONAL_LEADER* is 1 if an auditor is the number one auditor in an industry in terms of aggregated audit fees in a specific fiscal year, and 0 otherwise; *CITY_LEADER* is 1 if an office is the number one auditor in terms of aggregated client audit fees in an industry within that city in a specific fiscal year, and 0 otherwise; *OPSEG* is the number of operating segments reported in the Compustat segments database for the firm in $year_t$; *GEOSEG* is the number of geographic segments reported in the Compustat segments database for the firm in $year_t$; *SALESGROWTH* is the one-year growth rate of a firm's sales revenue; *SALESVOLATILITY* is the standard deviation of sales revenue, we use a rolling window and require a minimum of three years of data; *CFOVOLATILITY* is the standard deviation of *CFO*, we use a rolling window and require a minimum of three years of data to estimate; *CFO* is operating cash deflated by lagged total assets; *LOSS* is 1 if net income before extraordinary items is negative, and 0 otherwise; *BANKRUPTCY* is the Altman Z-score, which is a measure of the probability of bankruptcy, with a lower value indicating greater financial distress; *VOLATILITY* is a client's stock volatility and is the standard deviation of 12 monthly stock returns for the current fiscal year; and *MB* is the natural logarithm of the ratio of a client's market value of equity to its book value of equity in $year_t$.

[16] When we include the extensive list of control variables as per FY, the Big 4 effect could disappear due to the multicollinearity of these auditor-related variables with the *BIG4* variable. For example, the correlation between *BIG4* and *LOGOFFICE* in our separate analyses is approximately 0.4. Thus, what we learn from the matched sample findings is that they are not likely caused by the multicollinearity of the additional auditor controls with the *BIG4* variable as the matched samples balance and, hence, mitigate these multicollinearity effects to some extent.

[17] A backfitting algorithm simultaneously estimates the effects of nonlinear terms and the coefficients for the linear terms. This algorithm is available in the R and S-Plus packages using the function gam (generalized additive model).

In addition to using the RESET test, we generate scatter plots with both linear and local polynomial fits of each nonlinear variable. These scatter plots support the RESET test conclusions concerning the documented nonlinearities in each audit-quality proxy sample. We also run a bootstrap analysis on the RESET analysis drawing 10,000 samples of 5,000 observations, and estimate the RESET tests using the bootstrapped estimates. We find that client size, measured as *LOG_ASSETS* or *LOG_MKT*, is the only variable nonlinear to both the auditor choice and audit-quality proxy in each of the audit-quality analyses. However, there could be other forms of nonlinearity that are not detected by these tests.

Heckman Self-Selection Model

To examine whether the results are sensitive to specifications that consider the Heckman (1979) procedure, we model the self-selection of auditors following Chaney et al. (2004).[18] We estimate the model separately for each of the three audit-quality proxy samples and include the inverse Mills' ratio, estimated by year, in the respective second-stage pooled regressions, in the full and matched samples.[19] In line with Francis et al.'s (2010) critique of the Heckman (1979) model pertaining to satisfying exclusion restrictions, we estimate the inverse Mills ratio controlling for client size using total assets, total sales, market value, and various combinations of these variables. We continue to find evidence of the Big 4 effect in our full samples and do not find an effect for our matched samples using these alternative specifications. Current research is attempting to extend the literature and resolve the differences in estimating treatment effects using either the Heckman (1979) or the propensity-score approaches (Guo and Fraser 2010).

Audit Fees and Client Characteristics

Several studies use audit fees as a proxy to assess the potential superiority of Big 4 auditors, generally finding that the audit fees of the Big 4 clients are significantly larger than those of the non-Big 4 clients.[20] Among other factors, fees proxy for the level of service provided (e.g., Davis et al. 1993; Whisenant et al. 2003) and are negatively associated with levels of earnings management (e.g., Frankel al. 2002; Ashbaugh et al. 2003). However, the existence of a fee premium in itself does not imply higher audit quality, especially if the Big 4 auditors have more pricing power over their clients than do the non-Big 4 auditors (Francis 2004). Moreover, O'Keefe et al. (1994, 242) caution that "inferences about prices in such studies can be erroneous if the cross-sectional variations in auditor effort caused by differences in client characteristics are not adequately controlled."

As a sensitivity analysis, we apply our matching methodology to examine whether the Big 4 premium may be attributed to client characteristics. Using an audit fee model similar to that employed by Chaney et al. (2004), we perform the same set of analyses (untabulated) as with the prior three audit-quality proxies, using Compustat fee data from 2000 to 2006. We find that the

[18] Following Chaney et al. (2004), we model the self-selection of auditors, estimated separately for each year, as follows in order to obtain the inverse Mills ratio:

$$BIG4_i = \beta_0 + \beta_1 LOG\_ASSETS_i + \beta_2 ATURN_i + \beta_3 CURR_i + \beta_4 DTA_i + \beta_5 ROA_i + \beta_6 ROALOSS_i + \varepsilon_i,$$

where *BIG4* equals 1 if the client has a Big 4 Auditor, and 0 otherwise; *LOG_ASSETS* equals the natural logarithm of total assets; *ATURN* equals $\text{sales}_t/\text{total assets}_t$; *CURR* equals $\text{current assets}_t/\text{current liabilities}_t$; *DTA* equals $\text{total debt}_t/\text{total assets}_t$; *ROA* equals $\text{net income before extraordinary items}_t/\text{average total assets}_{t-1}$; and *ROALOSS* equals *ROA* multiplied by 1 if client has negative net income, and 0 otherwise.

[19] Other articles in this literature that apply the Heckman model to panel data in this manner include Khurana and Raman (2004), and Behn et al. (2008).

[20] Such studies include: Simunic (1980), Francis (1984), Francis and Stokes (1986), Chan et al. (1993), Craswell et al. (1995), DeFond et al. (2000), Ferguson et al. (2003), Whisenant et al. (2003), and Chaney et al. (2004). For an extensive review of this literature, see Hay et al. (2006).

coefficient on *BIG4* is positive and statistically significant at the 5 percent level for the full, the propensity-score matched, and the attribute-based matched samples.[21] Our results are consistent with the propensity-score matched fee results presented in Clatworthy et al. (2009), and could be a reflection of the following circumstances noted in Simunic (1980) and O'Keefe et al. (1994): (1) that Big 4 auditors provide a higher quantity of audit services; (2) that Big 4 auditors charge a higher unit price; and (3) it is possible that we still do not adequately control for all the relevant client characteristics that drive the quantity of audit services rendered. Our analyses are robust to the inclusion of a comprehensive list of client and auditor characteristics as in Francis and Yu (2009), and to different matching specifications.

IX. CONCLUSION

In this study, we examine whether differences in quality between Big 4 and non-Big 4 audit firms could be a reflection of client characteristics. Using matching models or controlling for an extensive list of client and auditor variables, we find that the treatment effects of Big 4 auditors are insignificantly different from those of non-Big 4 auditors with respect to discretionary accruals, the *ex ante* cost-of-equity capital, and analyst forecast accuracy.

We caution the reader that our findings must be interpreted with due regard to their methodological limitations. First, an inherent limitation of this approach is that we are unable to match on pre-treatment attributes. Second, the propensity-score and attribute-based matching models rely on the assumption that the effects of unobservables are not pertinent to estimating the treatment effects. Third, matching reflects a trade-off between identifying the treatment effects and the ability to generalize results to the full population. Fourth, we cannot ensure that we include all relevant client and auditor control variables. We cannot rule out the possibility that the foregoing limitations could introduce biases into our analyses.

We reemphasize that our study does not resolve the underlying question as to whether differences in audit-quality proxies between Big 4 and non-Big 4 auditors can be attributed to client characteristics, but rather that it only provides suggestive evidence. We hope that our results encourage other researchers to explore alternative methodologies that further disentangle client characteristics from audit-quality effects.

REFERENCES

Armstrong, C., A. Jagolinzer, and D. Larcker. 2010. Chief executive officer equity incentives and accounting irregularities. *Journal of Accounting Research* 48 (2): 225–271.

Ashbaugh, H., R. LaFond, and B. W. Mayhew. 2003. Do nonaudit services compromise auditor independence? Further evidence. *The Accounting Review* 78 (3): 611–639.

[21] Following Chaney et al. (2004), we model the determination of audit fees as follows:

$$LOGFEES_{i,t} = \beta_0 + \beta_1 BIG4_{i,t} + \beta_2 LOG\_ASSETS_{i,t} + \beta_3 ATURN_{i,t} + \beta_4 EXPORT_{i,t} + \beta_5 CURR_{i,t} + \beta_6 LEV_{i,t} + \beta_7 QUICK_{i,t} + \beta_8 ROA_{i,t} + \beta_9 ROA\_LOSS_{i,t} + \beta_{10} ABSEXTRA_{i,t} + Industry\_FE + Year\_FE + \varepsilon_{i,t},$$

where *LOGFEES* equals the natural logarithm of total audit fees in $year_t$; *BIG4* equals 1 if the client has a Big 4 auditor in $year_t$, and 0 otherwise; *LOG_ASSETS* equals the natural logarithm of total assets at the end of $year_t$; *ATURN* equals $sales_t$/total $assets_{t-1}$; *EXPORT* equals the sum of sales from foreign segments in $year_t$/total sales in $year_t$; *CURR* equals current $assets_t$/current $liabilities_t$; *LEV* equals (long-term $debt_t$ plus debt in current $liabilities_t$)/total $assets_{t-1}$; *QUICK* equals (current $assets_t$ minus $inventory_t$)/current $liabilities_t$; *ROA* equals net $income_t$/total $assets_{t-1}$; *ROA_LOSS* equals *ROA* multiplied by 1 if client has negative net $income_{t-1}$, and 0 otherwise; and *ABSEXTRA* is the absolute value of extraordinary $items_t$/total $assets_t$. We find that inferences are robust to the inclusion of nonaudit fees and interaction terms between *BIG4*, complexity, and foreign segment proxies.

Beatty, R. 1989. Auditor reputations and the pricing of initial public offerings. *The Accounting Review* 64 (4): 693–709.

Becker, C., M. DeFond, J. Jiambalvo, and K. R. Subramanyam. 1998. The effect of audit quality on earnings management. *Contemporary Accounting Research* 15 (1): 1–24.

Behn, B., J. H. Choi, and T. Kang. 2008. Audit quality and properties of analyst earnings forecasts. *The Accounting Review* 83 (2): 327–359.

Beneish, M. 1999. Incentives and penalties related to earnings overstatements that violate GAAP. *The Accounting Review* 74 (4): 425–457.

Butler, M., A. Leone, and M. Willenborg. 2004. An empirical analysis of auditor reporting and its association with abnormal accruals. *Journal of Accounting and Economics* 37 (2): 139–165.

CFA Institute Center. 2008. *June Survey Where Do You Stand? Results on Auditors*. Charlottesville, VA: CFA Institute.

———. 2009. DG MARKT: Consultation on control structures in audit firms and their consequences on the audit market. Working paper, CFA Institute.

Chan, P., M. Ezzamel, and D. Gwilliam. 1993. Determinants of audit fees for quoted UK companies. *Journal of Business Finance & Accounting* 20 (6): 765–786.

Chaney, P., D. Jeter, and L. Shivakumar. 2004. Self-selection of auditors and audit pricing in private firms. *The Accounting Review* 79 (1): 51–72.

Clatworthy, M., G. Makepeace, and M. Peel. 2009. Selection bias and the Big Four premium: New evidence using Heckman and matching models. *Accounting and Business Research* 39 (2): 139–166.

Cram, D., V. Karan, and I. Stuart. 2009. Three threats to validity of choice-based and matched sample studies in accounting research. *Contemporary Accounting Research* 26 (2): 477–516.

Craswell, A., J. Francis, and S. Taylor. 1995. Auditor brand name reputations and industry specializations. *Journal of Accounting and Economics* 20 (3): 297–322.

Davis, L. R., D. N. Ricchiute, and G. Trompeter. 1993. Audit effort, audit fees, and the provision of nonaudit services to audit clients. *The Accounting Review* 68 (1): 135–150.

DeAngelo, L. 1981. Auditor size and audit quality. *Journal of Accounting and Economics* 3 (3): 183–199.

Dechow, P., R. Sloan, and A. Sweeney. 1996. Causes and consequences of earnings manipulation: An analysis of firms subject to enforcement actions by the SEC. *Contemporary Accounting Research* 13 (1): 1–36.

DeFond, M., J. Francis, and T. Wong. 2000. Auditor industry specialization and market segmentation: Evidence from Hong Kong. *Auditing: A Journal of Practice & Theory* 19 (1): 49–66.

Dehejia, R., and S. Wahba. 1999. Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association* 94 (4): 1053–1062.

———, and ———. 2002. Propensity score-matching methods for nonexperimental causal studies. *The Review of Economics and Statistics* 84 (1): 151–161.

———. 2005. Practical propensity-score matching: A reply to Smith and Todd. *Journal of Econometrics* 125 (1–2): 355–364.

Dopuch, N., and D. Simunic. 1980. The nature of competition in the auditing profession: a descriptive and normative view. In *Regulation and the Accounting Profession*, 34, (2): edited by J. Buckley and F. Weston, 283–289. Belmont, CA: Lifetime Learning Publications.

Dow Jones News Service. 2002. Grant Thornton acquires Andersen partners in Houston. (June 13).

Easton, P. 2004. PE ratios, PEG ratios, and estimating the implied expected rate of return on equity capital. *The Accounting Review* 79 (1): 73–95.

Farber, D. 2005. Restoring trust after fraud: Does corporate governance matter? *The Accounting Review* 80 (2): 539–561.

Ferguson, A., J. Francis, and D. Stokes. 2003. The effects of firm-wide and office-level industry expertise on audit pricing. *The Accounting Review* 78 (2): 429–448.

Feroz, E., K. Park, and V. Pastena. 1991. The financial and market effects of the SEC's accounting and auditing enforcement releases. *Journal of Accounting Research* 29 (Supplement): 107–142.

Financial Accounting Standards Board (FASB). 1987. *Statement of Cash Flows*. Statement of Financial Accounting Standards No. 95. Norwalk, CT: FASB.

Francis, J. 1984. The effect of audit firm size on audit prices: A study of the Australian market. *Journal of Accounting and Economics* 6 (2): 133–151.

———, and D. Stokes. 1986. Audit prices, product differentiation, and scale economies: Further evidence from the Australian market. *Journal of Accounting Research* 24 (2): 383–393.

———, and J. Krishnan. 1999. Accounting accruals and auditor reporting conservatism. *Contemporary Accounting Research* 16 (1): 135–165.

———, E. Maydew, and H. Sparks. 1999. The role of Big 6 auditors in the credible reporting of accruals. *Auditing: A Journal of Practice & Theory* 18 (2): 17–34.

———. 2004. What do we know about audit quality? *The British Accounting Review* 36 (4): 345–368.

———, C. Lennox, and Z. Wang. 2010. Selection models in accounting research. Working paper, University of Missouri–Columbia, Hong Kong University of Science and Technology, and Nanyang Technological University.

———, and M. Yu. 2009. Big Four office size and audit quality. *The Accounting Review* 84 (5): 1521–1552.

Frankel, R., M. Johnson, and K. Nelson. 2002. The relation between auditors' fees for nonaudit services and earnings management. *The Accounting Review* 77 (Supplement): 71–105.

Geiger, M., and D. Rama. 2006. Audit firm size and going-concern reporting accuracy. *Accounting Horizons* 20 (1): 1–17.

Government Accountability Office (GAO). 2003. *Public Accounting Firms. Mandated Study on Consolidation and Competition.* Washington, D.C.: GAO.

———. 2008. *Audits of Public Companies. Continued Concentration in Audit Market for Large Public Companies does not Call for Immediate Action.* Washington, D.C.: GAO.

Guay, W., S. P. Kothari, and R. Watts. 1996. A market-based evaluation of discretionary-accrual models. *Journal of Accounting Research* 34 (Supplement): 83–115.

Guenther, D. A., and M. Willenborg. 1999. Capital gains tax rates and the cost of capital for small business: Evidence from the IPO market. *Journal of Financial Economics* 53 (3): 385–408.

Gul, F., K. J. Kim, and A. Qiu. 2009. Ownership concentration, foreign shareholding, audit quality, and stock price synchronicity: Evidence from China. *Journal of Financial Economics.* 95 (3): 425–442

Guo, S., and M. Fraser. 2010. *Propensity Score Analysis: Statistical Methods and Applications.* Thousand Oaks, CA: SAGE Publications, Inc.

Hastie, T., and R. Tibshirani. 1990. *Generalized Additive Models.* London, U.K.: Chapman and Hall.

Hay, D. C., W. R. Knechel, and N. Wong. 2006. Audit fees: A meta-analysis of the effect of supply and demand attributes. *Contemporary Accounting Research* 23 (1): 141–191.

Heckman, J. 1979. The sample bias as a specification error. *Econometrica* 47 (1): 153–162.

———, H. Ichimura, and P. Todd. 1997. Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The Review of Economic Studies* 64 (4): 605–654.

———, ———, and ———. 1998. Matching as an econometric evaluation estimator. *The Review of Economic Studies* 65 (2): 261–294.

———, and S. Navarro-Lozano. 2004. Using matching, instrumental variables, and control functions to estimate economic choice models. *The Review of Economics and Statistics* 86 (1): 30–57.

Hribar, P., and C. Nichols. 2007. The use of unsigned earnings quality measures in tests of earnings management. *Journal of Accounting Research* 45 (5): 1017–1053.

Krishnan, G. 2003. Audit quality and the pricing of discretionary accruals. *Auditing: A Journal of Practice & Theory* 22 (1): 109–126.

Khurana, I., and K. Raman. 2004. Litigation risk and the financial reporting credibility of Big 4 versus non-Big 4 audits: Evidence from Anglo-American countries. *The Accounting Review* 79 (2): 473–495.

Kothari, S. P., A. Leone, and C. Wasley. 2005. Performance matched discretionary accrual measures. *Journal of Accounting and Economics* 39 (1): 163–197.

Lang, M., and R. Lundholm. 1996. Corporate disclosure policy and analyst behavior. *The Accounting Review* 71 (4): 467–492.

Lennox, C. 1999. Are large auditors more accurate than small auditors? *Accounting and Business Research* 29 (3): 217–228.

Li, K., and N. Prabhala. 2007. Self-selection models in corporate finance. In *Handbook of Corporate Finance: Empirical Corporate Finance*, edited by B. E. Eckso, 37–86. Amsterdam, The Netherlands: Elsevier Science B.V.

Louis, H. 2005. Acquirers' abnormal returns and the non-Big 4 auditor clientele effect. *Journal of Accounting and Economics* 40 (1–3): 75–99.

Mitton, T. 2002. A cross-firm analysis of the impact of corporate governance on the East Asian financial crisis. *Journal of Financial Economics* 64 (2): 215–241.

O'Keefe, T. B., D. A. Simunic, and M. T. Stein. 1994. The production of audit services: Evidence from a major public accounting firm. *Journal of Accounting Research* 32 (2): 241–261.

Palmrose, Z. V. 1988. An analysis of auditor litigation and audit service quality. *The Accounting Review* 63 (1): 55–73.

Ramsey, J. 1969. Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society. Series A (General)* 31 (2): 350–371.

Rosenbaum, P., and D. Rubin. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70 (1): 41–55.

Simunic, D. 1980. The pricing of audit services: Theory and evidence. *Journal of Accounting Research* 18 (1): 161–190.

Smart, S., and C. Zutter. 2003. Control as a motivation for under pricing: A comparison of dual and single-class IPOs. *Journal of Financial Economics* 69 (1): 85–110.

Weber, J., and M. Willenborg. 2003. Do expert informational intermediaries add value? Evidence from auditors in microcap IPOs. *Journal of Accounting Research* 41 (4): 681–720.

Whisenant, S., S. Sankaraguruswamy, and K. Raghunandan. 2003. Evidence on the joint determination of audit and nonaudit fees. *Journal of Accounting Research* 41 (4): 721–744.

White, H.. 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48 (4): 817–838.

Zmijewski, M. 1984. Methodological issues related to the estimation of financial distress prediction models.. *Journal of Accounting Research* 22 (Supplement): 59–82.

THE ACCOUNTING REVIEW
Vol. 86, No. 1
2011
pp. 287–323

American Accounting Association
DOI: 10.2308/accr.00000016

The Role of the Internal Audit Function in the Disclosure of Material Weaknesses

Shu Lin
California State University, Fresno

Mina Pizzini
Southern Methodist University

Mark Vargus
Drexel University

Indranil R. Bardhan
The University of Texas at Dallas

ABSTRACT: This study investigates the role that a firm's internal audit function (IAF) plays in the disclosure of material weaknesses reported under Section 404 of the Sarbanes-Oxley Act of 2002 (U.S. Congress 2002). Using data from 214 firms, we examine the relation between material weakness (MW) disclosures and various IAF attributes and activities. Our results indicate that MW disclosures are negatively associated with the education level of the IAF and the extent to which the IAF incorporates quality assurance techniques into fieldwork, audits activities related to financial reporting, and monitors the remediation of previously identified control problems. The timing of Section 404 work and the nature of follow-up monitoring suggests that these aspects of IAF quality help prevent MWs from occurring. We find that MW disclosures are positively associated with the IAF practice of grading audit engagements and external-internal auditor coordination, suggesting that these activities increase the effectiveness of Section 404 compliance processes.

Keywords: *internal audit function; material weakness; corporate governance; internal control over financial reporting.*

Data Availability: *The Institute of Internal Auditors controls access to the Global Auditing Information Network (GAIN) survey data.*

We thank the Institute of Internal Auditors for access to the Global Auditing Information Network (GAIN) database. We thank Jeffrey Doyle, Weili Ge, and Sarah McVay for access to material weakness data reported in Doyle et al. (2007a). We also thank Larry Abbott, Barry Brian, Margaret Christ, Mark Salamasick, Douglas Ziegenfuss, Steven Kachelmeier (editor), two anonymous referees, the WWMA, and participants at the 2008 AAA Management Accounting Section Meeting, 2008 AAA Auditing Section Meeting, and the 2008 Lone Star Accounting Research Conference for their helpful comments and suggestions.

Editor's note: Accepted by Steven Kachelmeier.

Submitted: March 2008
Accepted: June 2010
Published Online: January 2011

I. INTRODUCTION

This study investigates the role that a firm's internal audit function (IAF) plays in the disclosure of material weaknesses reported under Section 404 of the Sarbanes-Oxley Act of 2002 (U.S. Congress 2002). Management is ultimately responsible for establishing and maintaining adequate internal control over financial reporting and for evaluating the effectiveness of financial reporting controls and procedures. However, "support for management in the discharge of these responsibilities is a legitimate role for internal auditors," as long as it does not impair internal audit objectivity (Institute of Internal Auditors [IIA] 2004, 3). External auditing standards have long acknowledged internal auditing as a potentially valuable resource in the financial reporting process (SAS No. 65, AICPA 1991; AS No. 2, PCAOB 2004; AS No. 5, PCAOB 2007a). Consistent with this reasoning, Bailey et al. (2003) find that companies view the IAF as a means of correcting perceived breakdowns in business reporting, internal control, and ethical behavior.

Despite the IAF's duties surrounding internal control over financial reporting (ICFR), few researchers have empirically investigated the IAF's role in the financial reporting process (Gramling et al. 2004). A notable exception is a recent study by Prawitt et al. (2009), which provides evidence that the IAF can improve reporting quality by mitigating potential weaknesses in incentive system design. Our study complements theirs, in that we examine the association between the IAF and ICFR through prevention and detection of material weaknesses. Accordingly, our study helps fill an important gap in the literature regarding the influence of the IAF on the quality of the financial reporting process.

Professional standards and prior research (AICPA 1991; IIA 2008; Prawitt et al. 2009) suggest that IAF quality encompasses specific attributes of the organizations and parties performing internal audit activities (e.g., competence of IAF personnel), as well as the nature and scope of activities performed (e.g., the extent to which IAF monitors remediation of previously identified control problems). We investigate the relation between these factors and the likelihood that a firm reports a material weakness (MW). For a material weakness to be disclosed, it must exist and be detected. It is not clear *ex ante* how our IAF quality measures will affect disclosures. For instance, more competent IAF personnel can help management establish stronger controls over financial reporting, and thus reduce the existence of control problems. Conversely, if a material weakness exists, more competent IAF personnel are more likely to detect it. We carefully consider the influence of various aspects of IAF quality on the existence, detection, and disclosure of MWs in developing our hypotheses.

We conduct our tests using data on 214 firms that provided detailed responses to the IIA's Global Auditing Information Network (GAIN) survey for 2003–2004. We identify 45 firms that disclosed at least one MW under SOX Section 404. Results indicate that the education level of IAF staff and the extent to which the IAF incorporates quality assurance techniques into fieldwork, audits activities related to financial reporting, and monitors the remediation of previously identified control problems are negatively associated with MW disclosures. The timing of Section 404 work and the nature of follow-up monitoring suggest that these aspects of IAF quality help prevent the existence of MWs. The IAF practices of grading audit engagements and coordinating with external auditors are both positively associated with MW disclosures. The positive relations suggest that these activities increase the effectiveness of Section 404 compliance processes and thereby increase the likelihood that extant MWs will be detected and disclosed. Together, our results have important implications for managers responsible for determining IAF staffing and the structure of IAF activities, external auditors who perform Section 404 work, and standard-setters who provide Section 404 guidance. Moreover, our evidence that the IAF affects the financial reporting process lends support to the requirement that NYSE-listed companies maintain an internal audit function (NYSE 2009).

This study makes several important contributions to the literature. First, we expand extant research on both internal auditing and ICFR by documenting associations between material weaknesses and various measures of IAF quality. Second, we identify specific internal auditor practices and procedures that are associated with both the detection and prevention of material weaknesses. Few archival studies directly link auditor practices to the prevention or detection of audit exceptions. Finally, we provide evidence that external auditors are more likely to detect material weaknesses when they coordinate their efforts with the IAF. While a significant branch of the internal auditing literature focuses on the relation between external auditors and the IAF, no study of which we are aware has examined whether IAF involvement in external audits increases the effectiveness of external audits.

Section II provides background information and reviews relevant literature, Section III describes our hypotheses, and Section IV describes the sample and model. Results are discussed in Section V, with concluding remarks in Section VI.

II. BACKGROUND AND LITERATURE REVIEW

Material Weakness Reporting

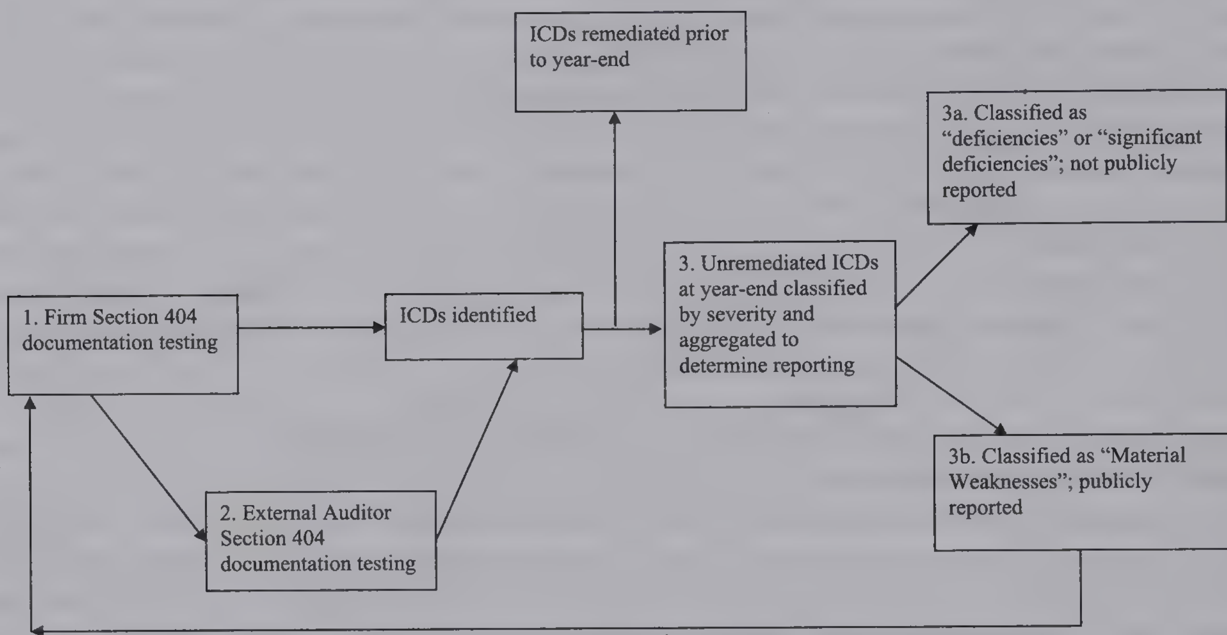
A material weakness is a deficiency, or combination of deficiencies, that result in a reasonable possibility that a company's controls will fail to prevent or detect a material misstatement of an account balance or disclosure (AS No. 5, PCAOB 2007a).[1] Section 404 of SOX mandates that managers evaluate internal control over financial reporting (ICFR) and present the results of their evaluation in financial statements filed on Form 10-K and Form 10-Q. The regulation, which became effective for accelerated filers for year-end dates beginning November 15, 2004, also requires external auditors to annually assess and state an opinion on ICFR. Accordingly, both management and external auditors are responsible for ensuring that material weaknesses are detected and disclosed under Section 404. Prior to the implementation of Section 404, Section 302 required management to evaluate and report on the effectiveness of ICFR; however, the external auditor was not required to opine on ICFR (U.S. Congress 2002). We exclude Section 302 material weaknesses from our analysis because these disclosures were subject to less regulation and allowed greater management discretion than were material weaknesses reported under SOX 404 (Ashbaugh-Skaife et al. 2007; Hoitash et al. 2009).

For a material weakness to be reported, it must exist, it must be detected, and it must be disclosed (Ashbaugh-Skaife et al. 2007). Figure 1 summarizes the sequence of Section 404 processes in a given year (Bedard and Graham 2011). Step 1 represents the firm's documentation and testing performed in support of management's evaluation of internal controls. Step 2 represents the external auditor's documentation and testing, which *must follow* the firm's work in a given area/account. At the end of the year, internal control deficiencies (ICDs) that have not been remediated are classified by severity and aggregated to determine whether they constitute a material weakness and, hence, must be publicly disclosed. The PCAOB and SEC direct external auditors and managers to evaluate the severity of each control deficiency to determine whether the deficiencies, individually or in combination, constitute a material weakness as of the date of management's assessment (AS No. 5, PCAOB 2007a; SEC 2007). Severity depends upon whether there is a reasonable possibility that the company's controls will fail to prevent or detect a misstatement and the magnitude of the potential misstatement. Multiple deficiencies that affect the same account or

---

[1] Although AS No. 2 was in effect during our sample period, we refer to AS No. 5, which supersedes AS No. 2, because we do not expect the differences in the two standards to affect our predictions regarding IAF quality (PCAOB 2004, 2007a). Two key differences are that AS No. 5 eliminates the redundant requirement that external auditors opine on both internal controls and on management's assessment of internal controls and replaces the words "more than remote" with the words "reasonable possibility" for defining what constitutes a material weakness.

FIGURE 1
Sequence of Section 404 Compliance Processes



disclosure may, in combination, constitute a material weakness, even though individually the deficiencies are less severe.

Gramling et al. (2004, 196) suggest that management's increased accountability for ICFR under SOX implies an expanded role for the IAF. Consistent with this reasoning, the IIA provides specific guidance regarding how the IAF can support management's SOX compliance (IIA 2004). The IIA advocates that the IAF independently evaluate management's testing and assessment processes, and management's basis for their assertions regarding the adequacy of internal controls. If control gaps are identified, then internal auditing should assess management's plans for correcting them and perform follow-up reviews. The IAF can also perform effectiveness testing for reliance by external auditors. Finally, the IIA advocates that the IAF act as coordinator between management and the external auditors and ensure that the results of ongoing internal audit activities are disclosed.

Prior Research on Auditor Quality

Practitioners and academics alike generally contend that the effectiveness of internal controls is increasing in IAF quality. However, direct empirical evidence of this relation is limited (e.g., AICPA 1991; Kinney 2000; Bailey et al. 2003; Gramling et al. 2004; PCAOB 2007a; Prawitt et al. 2009).[2] In an experimental setting, Schneider and Wilner (1990) find that, under certain conditions, internal auditing deters financial reporting irregularities. In an archival study of Australian firms, Davidson et al. (2005) find no significant relation between the presence (versus absence) of an IAF and earnings management. Prawitt et al. (2009) investigate the relation between IAF quality and earnings management using the GAIN database. Guided by external auditing standards, they develop a composite measure of IAF quality from proxies for IAF competence, IAF

[2] See Gramling et al. (2004) for a review of the literature on internal auditing.

objectivity, the amount of financial reporting work the IAF performs, and IAF size. They find evidence that IAF quality is associated with a lower level of earnings management. Their results suggest that higher-quality IAFs are more likely to deter managers from manipulating earnings and/or detect such manipulations and ensure that they are corrected prior to the issuance of financial statements.

Two recent studies of ICDs identified under Sections 302 and 404 suggest that the quality of the auditors performing Section 302 and Section 404 work is positively associated with the detection and disclosure of control problems. Using proprietary data on small accelerated filers, Bedard and Graham (2011) find that firms that use outside consultants for Section 404 compliance report higher levels of ICD severity. They posit that outside consultants have greater expertise in control assessments that improves the effectiveness of Section 404 processes and increases the likelihood of ICD detection and the appropriateness of ICD classification. In a study of non-accelerated filers, Bedard et al. (2009) find that the disclosure of Section 302 ICDs increases with the experience of the external audit firm in conducting Section 404 work, suggesting that audit firms leverage the knowledge gained from SOX 404 audits in performing SOX 302 work. They also find that more Section 302 MWs are reported in the fourth quarter, when there is greater external auditor involvement relative to the first three quarters. They contend that external audit firms' expertise, experience, and compliance processes for Section 302 are superior to that of client firms, and thus lead to increased detection and disclosure of ICDs at year-end.[3] While these studies provide useful archival evidence on the relation between auditor characteristics and the detection and disclosure of ICDs, neither directly addresses internal auditor quality.

Research on the effects of third-party monitoring on management reporting is also relevant to internal auditing (Prawitt et al. 2009). Hoitash et al. (2009) find that greater accounting and financial supervisory experience on the audit committee is associated with a lower likelihood of Section 404 MW disclosures. Studies of management forecasts and communications show that management is less biased when their bias is likely to be perceived by others (Schwartz and Young 2002; Rogers and Stocken 2005). Brown and Pinello (2007) provide evidence that the annual reporting process, which calls for an external audit, mitigates earnings management in year-end reports relative to that in interim reports. They attribute this result, at least in part, to increased scrutiny of year-end reports. Accordingly, if the internal auditor function is viewed as a third party that monitors managers' actions on a year-round basis, then improvements in IAF quality should strengthen deterrence and detection mechanisms.

III. HYPOTHESIS DEVELOPMENT

The overall relation between IAF quality and MW disclosures depends on how various IAF attributes and activities affect the existence, detection, and disclosure of MWs. If the effectiveness of ICFR is increasing in IAF quality, then IAF quality should be negatively associated with the existence of control problems and positively associated with both the detection and disclosure of extant control problems. Greater IAF quality deters managers from taking actions that compromise controls and encourages managers to put strong controls in place. Albrecht and Albrecht (2004) note that an effective control structure is probably the single most important step to eliminate (or minimize) the opportunity to commit fraud. A more capable IAF is also more likely to detect and correct minor control problems before they become severe enough to be considered material weaknesses. Decreasing the existence of control deficiencies unambiguously reduces the likelihood that a material weakness is reported.

---

[3] Competing disclosure incentives of management and the external auditor also explain the increased disclosure rate (Ashbaugh-Skaife et al. 2007; Asare et al. 2007).

If a material weakness already exists, then we expect that the improved detection capabilities of higher quality IAFs will increase the likelihood of disclosure. When a material weakness exists, the relation between IAF detection capability and disclosure depends upon (1) the likelihood that the external auditors will identify the problem during their year-end audit without the help of the IAF,[4] and (2) management's ability to correct the weakness prior to year-end, which is affected by the timing of IAF Section 404 procedures. If we assume the external auditors will not discover an existing MW, then IAF detection directly increases the likelihood of disclosure.[5] In this scenario, management can avoid disclosure only if they can correct the problem prior to year-end.[6] If we assume that the external auditors will detect all extant MWs at year-end, regardless of IAF activities, then IAF detection of a MW prior to year-end can potentially decrease the likelihood of disclosure by providing management with an opportunity to remediate the control weakness.

However, remediation requires that the IAF detect the problem early enough in the fiscal year for management to rectify the problem and that management deems the problem severe enough to take action. Ettredge et al. (2006) and PCAOB reviews of AS No. 2 (PCAOB 2005, 2007b) suggest that firms' Section 404 procedures were performed near or at year-end during the time period of our sample (2004–2006), leaving management little time for remediation. Moreover, Bedard and Graham (2011) report that management tends to underclassify the severity of control problems relative to external auditors, and thus may not recognize that a control problem is serious enough to warrant corrective action. Given this evidence, we conclude that it is unlikely that management will be able to successfully remediate the control problems in a timely manner. Accordingly, we expect that better IAF detection capabilities will lead to more MW disclosures, although we cannot rule out the possibility that enhanced IAF detection capabilities can reduce MW disclosures.

We draw on Prawitt et al. (2009) and professional guidance to develop measures of IAF quality (AICPA 1991; IIA 2008). These suggest that IAF quality measures encompass (1) competence, (2) objectivity, (3) relative investment in the IAF, and (4) the nature and scope of IAF activities. We group the first three measures together as IAF attributes because they address the characteristics of organizations performing internal audit activities, and thus are covered by the Attribute Standards of the *International Standards for the Professional Practice of Internal Auditing* (IIA 2008). The last construct, the nature and scope of IAF activities, captures IAF practices that are guided by Performance Standards (IIA 2008).[7]

IAF Quality Attributes

IIA Attribute Standards stipulate that internal auditors possess the knowledge, skills, and other competencies needed to effectively carry out their responsibilities (IIA 2008). To prevent and detect internal control irregularities, internal auditors must have a thorough understanding of company operations, processes, and procedures, and they must be able to design and implement tests to determine whether processes and procedures are working as intended (Clark et al. 1980).

---

[4] The literature is only beginning to address this complex issue of which party (firm or external auditor) is more likely to detect control deficiencies (Bedard et al. 2009; Bedard and Graham 2011). Therefore, we do not attempt to assess the likelihood that the external auditors would fail to detect a MW had it not been identified by the internal auditors.

[5] IIA guidance specifies that the IAF report all significant risk exposures and control issues to senior management and the board. Senior management is obligated under Section 404 to disclose a material weakness. It is unlikely that both management and the audit committee would knowingly violate SOX by failing to disclose a significant control problem to the external auditors.

[6] According to the SEC, "if management's evaluation process identifies material weaknesses, but all material weaknesses are remediated by the end of the fiscal year, management may conclude that internal control over financial reporting is effective at the end of the fiscal year" (SEC 2007, 10, footnote 20).

[7] We use the most recent version of the IIA Standards, issued October 2008, to guide our choice of quality measures. This version is very similar to earlier versions that were in place at the time the GAIN survey was conducted.

External auditing standards state that external auditors should consider professional certification, professional experience, and training in evaluating internal auditor competence (SAS No. 65, AICPA 1991). Consistent with this statement, prior research suggests that external auditors' evaluations of internal auditor competence are based upon professional certifications (Brown 1983) and the experience of IAF personnel (Messier and Schneider 1988).

IIA Standards require internal auditors to be independent and objective in performing their work (IIA 2008). An objective IAF is less likely to be influenced by management in evaluating controls and reporting internal control problems to the audit committee. Professional governance guidelines and standards suggest that the reporting relationship between the Chief Audit Executive (CAE) and audit committee is a key determinant of internal auditor objectivity (Gramling et al. 2004). Consistent with this premise, Bedard and Graham (2011) find that ICD severity levels are higher when the party performing the firm's Section 404 work reports directly to the audit committee rather than to management.

Practitioners and academics generally contend that management can improve IAF quality by increasing resources allocated to the IAF (e.g., Gramling et al. 2004; Ge and McVay 2005). Greater IAF resources enable management to hire and retain more competent personnel and improve the effectiveness of IAF consulting and assurance activities. In a descriptive analysis of MW disclosures, Ge and McVay (2005) find that poor internal control is usually related to an insufficient commitment of resources to accounting controls. Studies of the economic determinants of ICD disclosures find that firms with fewer resources available for internal controls are more likely to disclose ICDs (Ashbaugh-Skaife et al. 2007; Doyle et al. 2007a). Recent external auditing studies suggest that the greater resources available to large and mid-tier audit firms, relative to small audit firms, enable them to develop more effective SOX compliance processes that lead to greater ICD detection and disclosure (Bedard et al. 2009; Bedard and Graham 2011). Thus, we expect firms with greater IAF investment will implement stronger Section 404 procedures in support of management's evaluation of ICFR.

It is not clear *ex ante* how IAF quality attributes will affect MW disclosures. Competence, objectivity, and investment enhance quality and, thus, should reduce the likelihood that material weaknesses exist. Conversely, if a material weakness does exist, a higher quality IAF will be more likely to detect it. Given the bi-directional implications for material weakness disclosures, we present a nondirectional hypothesis for the three quality attributes.

H1: The likelihood that a firm reports a material weakness is significantly associated with IAF competence, IAF objectivity, and IAF investment.

Nature and Scope of IAF Activities

External auditing standards contend that external auditors evaluate the nature, timing, and extent of IAF fieldwork in audit planning and determining whether to rely on the work of internal auditors. AICPA and IIA guidance suggest that the following factors are relevant to the financial reporting process: (1) use of fieldwork quality assurance techniques, (2) inclusion of financial reporting processes in audit scope, (3) communication of grades or summary opinions on control effectiveness, (4) follow-up of previously identified control problems, and (5) coordination with external auditors (AICPA 1991; IIA 2008).

The quality of audit fieldwork performed is critical to the detection of internal control deficiencies. Schneider (1984, 1985) and Brown and Karan (1986) find that external auditors place more emphasis on the quality of work performance than either competence or objectivity in evaluating the IAF. Quality assurance (QA) techniques help ensure that IAF fieldwork is effective and appropriate. QA activities include direct supervision, independent working paper review, solicitation of audit client feedback, peer review by fellow staff members, and the use of working paper checklists.

As Prawitt et al. (2009, 1261) note, firms have broad latitude in setting the scope of the IAF, and its focus can vary widely across organizations. IAF responsibilities generally encompass a range of control activities not directly related to financial reporting, such as operational audits, systems audits, and internal consulting projects. Since ICFR forms only a subset of a firm's overall internal controls, external auditors must evaluate the extent to which IAF activities are relevant to the financial statement audit in determining whether they can rely on IAF work (SAS No. 65; AICPA 1991). We expect that greater IAF attention to financial reporting will affect both the existence and detection of material weaknesses.

Internal auditing standards direct internal auditors to document and communicate engagement results to management, and, where appropriate, include their overall opinion (IIA 2008). A grade or rating is a succinct means of conveying an opinion on the risk posed by the unit or functional area audited, and the IIA provides specific guidance on the practice of rating internal controls (IIA 2009a). A recent study by PricewaterhouseCoopers (PwC 2006) reports an increase in the prevalence of grading and concludes that grading is now generally considered to be a best practice. Grading motivates managers to put adequate controls in place to avoid a poor grade and facilitates design and implementation of monitoring systems. Both help prevent the existence of control problems. Grading also promotes rapid assessment of control risks by management, audit committees, and external auditors (PwC 2006), and thus can help management and external auditors implement the risk-based approach to Section 404 compliance recommended by professional guidance (PCAOB 2007a; SEC 2007). As an attention-directing tool, grading improves an auditor's assessment of the risk of financial misstatements and thereby facilitates appropriate allocation of audit resources to the evaluation and testing of relevant company-level controls (Hogan and Wilkins 2008; PCAOB 2007a; Wright and Ashton 1989). Accordingly, we expect IAF grading to improve the effectiveness of both internal and external auditors' Section 404 audit procedures, thus increasing the likelihood that extant material weaknesses are detected.

As with the IAF quality attributes, it is not clear *ex ante* how using QA techniques, focusing on financial reporting controls, and grading internal controls will affect the likelihood that a material weakness is reported. These monitoring activities can prompt managers to take preventative action in anticipation of the IAF's rigorous review (Brown and Pinello 2007). Conversely, such activities increase the likelihood that the IAF discovers and discloses any existing control problems. Further, IAF grading can increase the effectiveness of both internal and external Section 404 compliance processes, leading to greater MW detection. Due to the bi-directional implications for material weakness disclosures, we test a nondirectional hypothesis.

H2: The likelihood that a firm reports a material weakness is associated with: (1) the extent to which the IAF uses quality assurance techniques, (2) the extent to which IAF activities address financial reporting activities, and (3) whether the IAF grades audit engagements.

IIA performance standards and SOX guidance require the Chief Audit Executive (CAE) to establish and maintain a system to monitor the disposition of previously identified control problems (IIA 2008, 2002). Follow-up procedures can reduce the likelihood that a material weakness exists at year-end and, hence, must be publicly disclosed. First, follow-up procedures provide the impetus for management to correct less severe control problems, thereby preventing them from becoming material weaknesses. Second, if the firm or its external auditors detect an existing material weakness at any time prior to year-end, then management can avoid disclosing it if they resolve the problem prior to year-end. Hence, IAF follow-up procedures can prompt management to remediate identified weaknesses in a timely manner. Accordingly, we predict that follow-up procedures will be negatively associated with MW disclosures.

H3: Firms whose IAFs follow-up on control problems are less likely to report material weaknesses.

Both internal and external auditing standards encourage auditors to coordinate efforts related to an integrated audit under AS No. 5 (PCAOB 2007a; IIA 2008). Methods of coordination include conducting joint risk or planning sessions, performing audits of specific processes or locations, and loaning IAF staff to the external auditor. As firms' "in-house" experts on internal control, internal auditors possess company-specific knowledge about controls, operations, and the financial reporting process that can aid external auditors in implementing the top-down, risk-based approach recommended in AS No. 5 (PCAOB 2007a). Coordination should improve an external auditor's assessment of the risk of misstatements, thus facilitating appropriate allocation of audit resources to evaluation and testing of relevant controls (Hogan and Wilkins 2008; PCAOB 2007a). Reliance on the work performed by the IAF, both independently and under the direction of the external auditors, can improve the effectiveness of Section 404 documentation and testing.[8] According to the PCAOB (2007b, 8–9), "an auditor who appropriately uses the work of others can achieve the objectives of the audit while not duplicating effort in lower-risk areas, and also is better able to focus his or her own efforts on higher-risk controls." Moreover, to the extent that coordination reduces time pressure on the external auditor, audit effectiveness should increase with coordination (McDaniel 1990). Given the value the IAF can add to the external auditor's Section 404 processes, we propose that internal-external auditor coordination will enhance the overall effectiveness of the external auditor's Section 404 process, thereby leading to greater detection and disclosure of MWs.[9]

Our prediction assumes that most of the external auditor's Section 404 documentation and testing takes place at year-end, giving management little opportunity to correct any control problems detected by the external auditors. While we recognize that coordination can and does occur throughout the year, we expect that much of the benefit of that coordination is realized at year-end, when external auditors conduct their most rigorous reviews of the financial statements and the reporting process (Frankel et al. 2002; Brown and Pinello 2007; Bedard et al. 2009). In particular, the rapid enactment of Section 404 combined with delays in managements' completion of SOX procedures left external auditors little choice but to perform most Section 404 work at year-end for the time period of our sample of MWs (PCAOB 2005; Ettredge et al. 2006). The benefits of IAF coordination and the year-end timing of external auditors' Section 404 processes lead to the following hypothesis:

H4: Firms in which IAFs coordinate with external auditors are more likely to report material weaknesses.

IV. RESEARCH DESIGN

Data and Sample Selection

Data for this study come from multiple sources. We use firm-level data collected by the IIA through their GAIN survey for the years 2003 and 2004. The GAIN database consists of CAEs' responses to a comprehensive survey designed to measure various aspects of an organization's

---

[8] The IAF must be of sufficient quality, i.e., it must be objective and competent for external auditors to rely on their work. Thus, even though the IAF are employees of the firm being audited, we expect that the IAF will be objective in assessing controls and reporting control problems to the external auditors.

[9] External auditor reliance on the IAF is increasing in IAF quality (e.g., Schneider 1984, 1985; Maletta 1993), and as previously discussed, it is not *a priori* clear how IAF quality affects the likelihood of MW disclosure. Our tests of the relation between IAF coordination and MW disclosure include explicit measures of IAF quality, and thereby control for the effect of IAF quality on disclosure.

internal audit activities. The annual survey captures information on several topics, including internal audit department costs, oversight and reporting responsibilities, audit committee, audit life-cycles, and financial measures.[10] Next, we collect firm data on MW disclosures from EDGAR, firm financial data from Compustat, stock return data from CRSP, audit fee and restatement data from Audit Analytics, and institutional ownership data from CDA/Spectrum.

Our initial sample contained 1,356 responses from 935 GAIN survey respondents collected in years 2003 and 2004. Since firm names are not reported in the survey data, we employ a matching algorithm to identify firms based on reported SIC code, total assets, revenues, and number of employees. Next, we eliminate: (1) 687 firms for which data were unavailable on Compustat, (2) six firms with American Depositary Receipts or missing CRSP data, (3) six firms with missing 10-K data after August 2002, (4) six firms missing stock information from both Compustat and CRSP, and (5) eight firms missing necessary GAIN data, resulting in a sample of 329 firm-year observations from 222 firms.[11] For firms that report survey data for both 2003 and 2004, we use only the data from the 2004 survey year, which yields our final sample of 222 firms. Finally, we eliminate five firms that only report material weaknesses under SOX Section 302, and three firms that are non-accelerated filers and therefore not subject to SOX Section 404 reporting requirements.[12] The final sample consists of 214 firms (Table 1).

IAF Quality as a Determinant of Material Weaknesses

We model the probability that a firm reports a material weakness as a function of IAF attributes, the nature and scope of IAF activities, and a set of control variables (Equation (1)). We estimate this model using a logit regression.

$$\begin{aligned} Prob(MW=1) = {} & \beta_0 + \beta_1 EXPERIENCE + \beta_2 EDUCATION + \beta_3 CERTIFICATION \\ & + \beta_4 TRAINING + \beta_5 CAEAC + \beta_6 CAEOFFICER + \beta_7 IASIZE \\ & + \beta_8 FIELDWORKQA + \beta_9 IAGRADE + \beta_{10} FINANCIALFOCUS \\ & + \beta_{11} FOLLOWUP + \beta_{12} COORDINATION + \beta_{13} SEGMENTS \\ & + \beta_{14} FOREIGNTRANSACTIONS + \beta_{15} M\&A + \beta_{16} RESTRUCTURE \\ & + \beta_{17} SALESGROWTH + \beta_{18} INVENTORY + \beta_{19} MARKETVALUE + \beta_{20} LOSS \\ & + \beta_{21} CFO + \beta_{22} SHUMWAY + \beta_{23} AGE + \beta_{24} AUDITORSPECIALIST \\ & + \beta_{25} BLUERIBBONAC + \beta_{26} ISO9000 + \beta_{27} RESTATEMENT \\ & + \beta_{28} AUDITORRESIGN + \beta_{29} INSTITUTIONALOWNERSHIP \\ & + \beta_{30} REGULATEDINDUSTRY + \varepsilon \end{aligned} \tag{1}$$

[10] The data were subject to various validation checks, including validation measures built into the questionnaire and manual procedures and reasonableness tests applied after the data had been collected. The GAIN database covers a wide range of institutions including publicly traded companies, private companies, educational institutions, divisions within companies, and governmental institutions. More information can be found at the GAIN website: http://www.theiia.org/guidance/benchmarking/gain/.

[11] We eliminate firms with missing values for *IASIZE*, *EXPERIENCE*, *EDUCATION*, and *CERTIFICATIONS*, and set missing values to zero for *TRAINING*, *OBJECTIVITY*, *FIELDWORK QA*, *FOLLOWUP*, *IAGRADE*, *FINANCIALFOCUS*, and *COORDINATION*. There are no missing values for *CAEOFFICER*, *BLUERIBBONAC*, and *ISO9000*. In a robustness check, we remove the firms with missing values in any IAF variables and obtain similar results.

[12] Inclusion of the five firms that report material weaknesses under Section 302 does not materially affect our results or conclusions. In untabulated robustness tests, the coefficient on *FINANCIALFOCUS* retains its hypothesized sign, but becomes insignificant at the 10 percent level; coefficients on the other IAF quality measures have the same signs and similar significance levels.

TABLE 1

Sample Selection and Distribution by Year

Panel A: Sample Selection

	Firm-Year Observations	Firms
Firm-year observations obtained from GAIN database for survey years 2003 and 2004[a]	1,356	935
Less:		
Those not covered by Compustat	(990)	(687)
American Depositary Receipts	(10)	(6)
Those missing SEC filings after August 2002	(6)	(6)
Those missing stock information from both Compustat and CRSP	(10)	(6)
Those missing IAF data from GAIN	(11)	(8)
	329	222
Elimination of 2003 firm-years if data from 2004 survey-year data are available[b]	(107)	—
Elimination of firms that only report material weakness(es) under SOX Section 302	(5)	(5)
Elimination of firms that are non-accelerated filers (with market value less than $75 million) and therefore are not subject to SOX Section 404	(3)	(3)
	214	214

Panel B: Distribution of IAF Data by Fiscal Years

Fiscal Year	n	%
2000	6	2.80
2001	42	19.63
2002	57	26.64
2003	97	45.32
2004	12	5.61
Total	214	100.00

[a] A survey collected in 2003 or 2004 may describe earlier fiscal years. For example, survey information collected in 2003 may pertain to 2002 or 2001.

[b] We retained the survey data that pre-dates the year in which a material weakness is first disclosed. In the final sample, all but two firms' IAF data describe firm characteristics that had been in existence at least one year prior to the year in which a material weakness had been disclosed. The empirical results remain similar after excluding the two firms from the sample.

where *MW* is an indicator variable that is equal to 1 if the firm disclosed a material weakness in internal control, and 0 otherwise. All model variables are defined in Table 2.

Material Weakness Firms

To construct the dependent variable, we examine 10-Ks and 10-Qs for the 214 sample firms in the EDGAR database and identify MW disclosures during the period November 15, 2004 to December 2006.[13] We identify 47 firms that disclosed at least one material weakness during this time frame. We remove two firms due to missing data, leaving 45 firms that report at least one material weakness.

[13] November 15, 2004 coincides with the date that Section 404 became operational for accelerated filers.

TABLE 2
Variable Definitions

Internal Control over Financial Reporting	
MW	= An indicator variable that equals 1 if the firm disclosed a material weakness under Section 302 or Section 404, and 0 otherwise.
IAF Attributes[a]	
EXPERIENCE	= Average number of years of auditing experience (internal and external) of the audit staff (B6a and B6b).
EDUCATION	= Average of the number of years of undergraduate and graduate education of the audit staff, based on highest degree achieved. Associate, Bachelor, Master, and Ph.D. degrees are assumed to take 2, 4, 6, and 8 years of study, respectively (B5).
CERTIFICATIONS	= Percentage of professional staff members with one or more audit certifications (B8).
TRAINING	= Annual hours of training per internal auditor (B15b).
CAEAC	= An average of eight 0-1 survey items that measures the amount of relevant internal control information the CAE reviews with the audit committee. Four items address the control environment: risk assessment system, overall assessment of corporate control environment, assessment of control environment by major subsidiary of operating entity, coordination of internal auditing with external auditor's plan. Four items address the IAF: significant findings, audits performed, fraud-conflicts of interest; and results of monitoring programs concerning compliance with law, code of conduct and ethics (C9, C10).
CAEOFFICER	= An indicator variable that equals 1 if the CAE position is an officer of the firm (E7).
IASIZE	= The total annual operating costs of the IAF (B1g) divided by the firm assets (A2), multiplied by 100.
IAF Activities[a]	
FIELDWORKQA	= An average of seven survey items coded from 0 to 2 (0: Never, 1: Occasionally, and 2: Regularly) that measures the extent to which various quality assurance (QA) techniques are used in fieldwork. The QA techniques are: direct supervision; independent working paper review; audit client feedback; peer review by fellow staff members, working paper checklists, and ticklists; management participation; and other (H4).
IAGRADE	= An indicator variable that equals 1 if the final internal audit report includes a grade or score as determined by the results of the audit (H12).
FINANCIALFOCUS	= An average of six survey items that measure how frequently the IAF performs audits of various financial activities (0: Never, 1: Occasionally, and 2: Regularly). The survey classifies the following five activities as financial: adequacy of internal accounting controls; accuracy, reliability, and completeness of financial records; usefulness of financial reporting for management control and decision making; impact of changes in accounting rules or regulations; interim quarterly financial results reported externally (G4). We also include the frequency of GAAP compliance audits (G3e).
FOLLOWUP	= An indicator variable that equals 1 if there is a formal follow-up procedure to test the implementation of corrective action (H11).
COORDINATION	= An indicator variable that equals 1 if internal audit coordinated audit services the external auditors (F9a).

(continued on next page)

TABLE 2 (continued)

Internal Control over Financial Reporting Control Variables	
SEGMENTS	= Natural log of the sum of the number of operating and geographic segments in 2003 reported by Compustat Segment file.
FOREIGNTRANSACTIONS	= An indicator variable that equals 1 if the firm reports a non-zero value for foreign currency translations (#150) in 2003.
M&A	= Proportion of years from 2001 to 2005 that a firm was involved in a merger or acquisition (Compustat AFNT #1).
RESTRUCTURE	= Proportion of years from 2001 to 2005 that a firm was involved in restructuring. We identify a firm undergoing a restructuring if any of the following Compustat data items are non-zero: #376, #377, #378, or #379.
SALESGROWTH	= Proportion of years from 2001 to 2005 that a firm's annual industry-adjusted sales growth (#12) falls into the top quintile.
INVENTORY	= Average inventory to total assets (#3/#6) from 2001 to 2005.
MARKETVALUE	= Natural log of the market value of equity (#199 × #25) in 2003.
LOSS	= Proportion of years from 2001 to 2005 that a firm reports negative earnings before extraordinary items (#18).
CFO	= Average cash flows from operations to total assets (#308/#6) from 2001 to 2005.
SHUMWAY	= The probability of bankruptcy as predicted by Shumway's (2001) default hazard model.
AGE	= Natural log of the number of years the firm has Compustat data since 1980.
AUDITORSPECIALIST	= An indicator variable that equals 1 if the firm was audited by an industry specialist auditor. We define industry specialist as an auditor that collects the greatest percentage of its audit fees in the client's industry.
BLUERIBBONAC	= A score with a range of 0 to 1 measuring the extent to which the firm has implemented the Blue Ribbon Audit Committee recommendations. The score is an average of ten Blue Ribbon Committee recommendations coded 0/1: all members meet new criteria of independence; at least three independent outside directors; all members possess core skills including finance literacy; committee's charter is reassessed annually; proxy report states committee fulfills its charter; committee is accountable for auditor relations; outside auditors disclose all consulting assignments; auditors discuss adequacy of company accounting; 10-K's MD&A discloses financial statement reviews and discussions; and auditors review quarterly reports and 10-Q before release.
ISO9000	= An indicator variable that equals 1 if the firm is involved in ISO 9000.
RESTATEMENT	= An indicator variable that equals 1 if the firm had a restatement from 2001 to 2005.
AUDITORRESIGN	= An indicator variable that equals 1 if the firm experienced an auditor resignation during 2001 to 2005.
INSTITUTIONALOWNERSHIP	= The percentage of institutional ownership as of December 31, 2003.
REGULATEDINDUSTRY	= An indicator variable that equals 1 if the firm is in financial services or utility industry.

[a] Letters and numbers in parentheses for IAF attributes and activities refer to item codes in 2004 GAIN Survey.

IAF Attributes

IAF quality attributes include the competence and objectivity of internal audit personnel and the amount firms invest in the IAF. In measuring these attributes, we generally follow professional guidance and Prawitt et al. (2009), who also use the GAIN data to measure IAF quality. To measure IAF competence, we use four variables, *EXPERIENCE*, *EDUCATION*, *CERTIFICATION*, and *TRAINING* (SAS No. 65, AICPA 1991; Prawitt et al. 2009). *EXPERIENCE* is defined as the average number of years of internal and external auditing experience of the audit staff; *EDUCATION* is the average number of years of undergraduate and graduate education; *CERTIFICATION* is the percentage of professional staff members with one or more audit certifications; and *TRAINING* is the average annual hours of training per staff member.[14] IIA standards and guidelines for objectivity recommend that the CAE regularly communicate the results of ongoing internal audit activities to the audit committee (Practice Advisory 1110-1, IIA 2009b, 1), and that the CAE report to a level within the organization "that allows the internal audit activity to fulfill its responsibilities" (Attribute Standard 1110, IIA 2008, 2). We capture CAE-audit committee communication with the variable *CAEAC*, which measures the amount of control-related information the CAE reviews with the audit committee. Examples of such information include assessments of the control environment in total and by major subsidiary, areas audited, significant findings, fraud, and ethics compliance. A second proxy for objectivity, *CAEOFFICER*, equals 1 if the CAE position is an officer of the firm, and 0 otherwise. We expect that a CAE who is an officer will have greater support from the board and senior management, which in turn should assist the IAF in performing work free from interference (Practice Advisory 1110-1, IIA 2009b).[15] We measure IAF investment, *IASIZE,* as the total IAF annual operating costs divided by total assets, based on a measure developed by Prawitt et al. (2009).[16]

Nature and Scope of IAF Activities

The nature and scope of IAF activities investigated includes: (1) use of quality assurance techniques in fieldwork, (2) inclusion of financial reporting processes in audit scope, (3) inclusion of grades or summary opinions on control effectiveness in audit reports, (4) follow-up of previously identified control problems, and (5) coordination with external auditors. We measure the extent to which the IAF uses quality assurance techniques, *FIELDWORKQA*, with a summative score of variables that measure the frequency with which various QA techniques are used. *FINANCIALFOCUS* is an index compiled from several items that measure the extent to which the IAF audits various activities related to financial reporting.[17] The indicator variable *IAGRADE* equals 1 if the final internal audit report includes a "grade" or "score" for areas reviewed. The indicator variable *FOLLOWUP* equals 1 if internal auditors formally monitor the resolution of

---

[14] Our measures for *EXPERIENCE* and *CERTIFICATIONS* differ slightly from Prawitt et al. (2009), while our training measure is the same. Prawitt et al. (2009) include only internal audit experience in their experience measure; we include both internal and external audit experience because external auditing experience is also valuable in conducting internal audits. Prawitt et al. (2009) calculate *CERTIFICATIONS* with two different survey items that measure separately the number of CIA and CPA designations. This approach double-counts individuals who have both designations. To avoid double-counting, we use a single survey question that asks for the "total number of professional staff with one or more audit certifications."

[15] Prawitt et al. (2009) measure objectivity with a single dichotomous variable indicating whether the CAE reports to the audit committee. Robustness tests using this variable give materially similar results in our sample.

[16] Prawitt et al. (2009) also scale IAF operating costs by assets. However, they then subtract the average level of industry investment (computed using all GAIN firms) from this amount and convert this measure to a dichotomous variable by assigning a value of 1 to firms whose investment equals or exceeds the average level of investment for that firm's particular industry. We use Prawitt et al.'s (2009) dichotomous measure in robustness tests and it does not affect the signs or significance of our IAF measures.

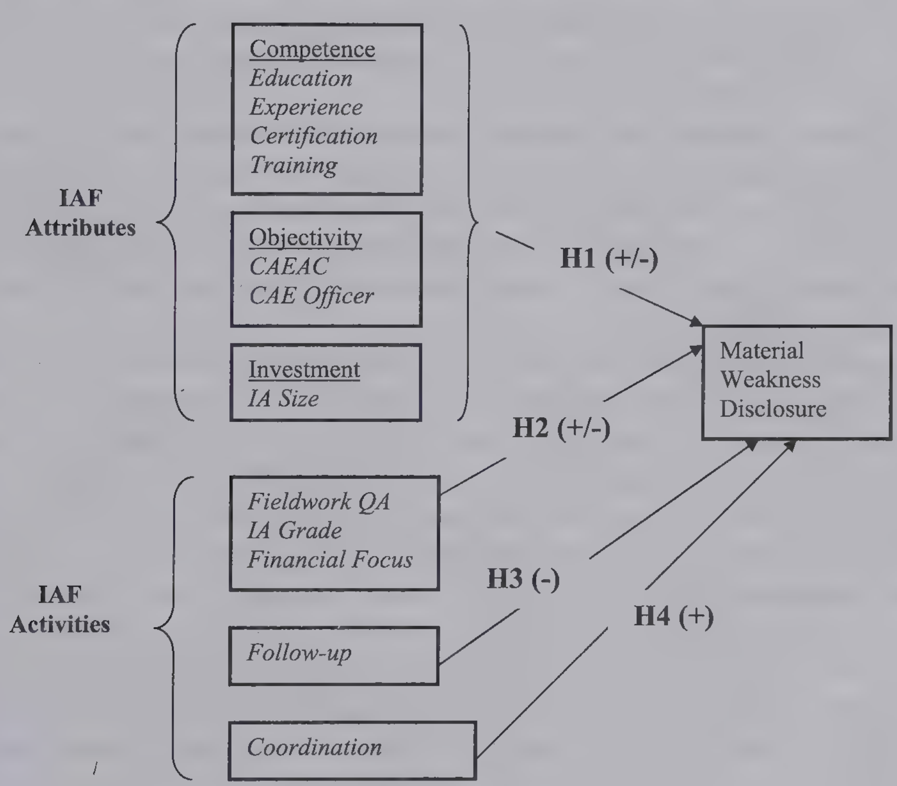
[17] Prawitt et al. (2009) measure *FINANCIALFOCUS* with a survey item that gives the percentage of time the IAF spends performing financial audit work. In our sample, there are too many missing values (78 out of 214) for this item.

previously identified control problems, and 0 otherwise. The indicator variable *COORDINATION* is coded 1 if internal auditors coordinate services with external auditors, and 0 otherwise. The survey cites the following coordination methods: loaning staff to external auditors; performing complete or partial audits of specific locations, products, or functions; conducting joint annual planning sessions; and conducting joint risk or control sessions. Figure 2 summarizes hypothesized associations between the likelihood of a MW disclosure and variables measuring IAF attributes.

Control Variables

In Equation (1) we use the following variables to control for several firm-specific factors that prior research has shown to be correlated with the likelihood that a MW is disclosed: (1) *SEGMENTS* and *FOREIGNTRANSACTIONS* proxy for internal control risks that stem from organizational complexity; (2) *RESTRUCTURE*, *M&A*, and *SALESGROWTH* proxy for internal control risk associated with rapid organizational change; (3) *INVENTORY* captures internal control risk associated with inventory and product obsolescence costs; (4) *MARKETVALUE* controls for the level of resources available to invest in internal controls; (5) *LOSS*, *CFO*, and *SHUMWAY* measure financial distress and reflect a firm's diminished ability to adequately invest in internal controls;[18]

FIGURE 2
Hypothesized Relations between IAF Attributes and Activities and the Likelihood That a Firm Discloses a Material Weakness



[18] We also used Altman's Z-score in place of the Shumway score to evaluate the sensitivity of the model. The signs and significance levels of IAF variables remain unchanged in our full model.

(6) *AGE* measures firm age; (7) *AUDITORSPECIALIST* is a dichotomous variable indicating whether the external auditor is an industry specialist; and (8) *BLUERIBBONAC* controls for the extent to which firms have implemented Blue Ribbon Committee recommendations (Carcello et al. 2005; Krishnan 2005; Ashbaugh-Skaife et al. 2007; Doyle et al. 2007a; Stephens 2009). We include *ISO9000* to control for firm certification by the International Standards Organization (ISO) because ISO implementation may potentially improve the internal controls associated with organizational business practices. The variables *RESTATEMENT* and *AUDITORRESIGN* control for existing internal control problems associated with earnings restatements and auditor resignations, respectively, during the 2001–2005 period. Finally, *INSTITUTIONALOWNERSHIP* and *REGULATEDINDUSTRY* proxy for external monitoring mechanisms that may be associated with the frequency of material weakness disclosures. Variable definitions are summarized in Table 2.

V. RESULTS

Descriptive Statistics

Table 3 provides descriptive statistics for the firms in our sample, along with comparative statistics for firms in the Compustat universe. Panel A of Table 3, which presents the distribution of sample firms by industry, indicates that the sample contains a high concentration of firms in regulated industries. Wilcoxon tests, reported in the extreme right column of Panel B, indicate that our sample firms are significantly larger, older, and more profitable than those in the Compustat universe. Prior research suggests that firms in regulated industries, and those that are older, larger, and more profitable, make greater investments in the IAF (Wallace and Kreutzfeldt 1991; Carcello et al. 2005; Goodwin and Kent 2006). The industry, size, and profitability comparisons are consistent with the propensity for GAIN survey respondents to represent firms that have relatively large internal audit functions. To the extent that firms with large internal audit functions are more prone to respond to the GAIN survey, our tests are potentially biased against detecting associations between IAF quality and material weaknesses.

Panel C of Table 3 compares the industry distribution of reported material weaknesses in our study sample to the distribution of firms reported in Doyle et al. (2007a). Approximately 21 percent of our sample firms disclosed *at least* one material weakness during the period November 2004 through December 2006. This is comparable to Doyle et al. (2007b), who find that 17 percent of their sample firms report material weaknesses during the period from August 2002 through October 2005.

Table 4 presents descriptive statistics for the independent variables in our model for the overall sample, partitioned on MW disclosures. Univariate tests of differences between the partitions indicate that material weakness firms are more likely to have internal auditors that issue grades in their audit reports (64 percent versus 42 percent; $p = 0.019$), and have fewer financial reporting activities audited by their internal auditors. Consistent with prior research, material weakness firms have significantly higher incidences of restructuring activities and losses, lower cash flows from operations, and higher bankruptcy risk. Univariate tests also show that a significantly greater proportion of material weakness firms report foreign currency adjustments and auditor resignations. Finally, the material weakness firms in our sample tend to be smaller and less likely to have *ISO9000* certifications. Overall, these results suggest that material weakness disclosures are influenced by differences in some IAF attributes in addition to previously documented firm characteristics.

Table 5 provides Spearman and Pearson correlation coefficients for the independent variables in our model. Since IAF measures attempt to capture the same underlying construct, quality, we find some significant correlations between IAF attributes and activities (Table 5, Panel A). In particular, two measures of competence, *CERTIFICATIONS* and *EXPERIENCE,* are highly correlated ($r = 0.39$). Also, *FINANCIALFOCUS* is significantly correlated with *CAEAC* ($r = 0.34$),

TABLE 3

Comparison of Sample Firms to Compustat Universe and Doyle et al. (2007a) Sample

Panel A: Distribution by Industry

Industry[a]	n	Sample (n = 214) %	2003 Compustat (n = 7,019) %
Banks and Insurance	38	17.76	20.80
Utilities	38	17.76	2.55
Durable Manufacturers	33	15.42	15.55
Computers	22	10.28	15.07
Textiles, Printing, and Publishing	20	9.35	3.36
Transportation	15	7.01	6.15
Retail	15	7.01	7.81
Services	8	3.74	8.32
Chemicals	9	4.21	2.14
Refining and Extractive	4	1.87	3.46
Food	6	2.80	1.97
Drug and Medical Equipment	4	1.87	9.06
Mining and Construction	1	0.47	2.45
Miscellaneous	1	0.47	1.01
Agriculture	—	—	0.30
Total	214	100.00	100.00

Panel B: Distribution by Firm Characteristics

	Sample (n = 214)		2003 Compustat (excluding sample firms)[b]		
	Mean	Median	Mean	Median	Median Comparison[c]
Market Value (millions)	14,542.16	4683.13	2,474.69	180.46	<0.001
Total Assets (millions)	25,601.42	7,835.65	6,208.13	226.90	<0.001
Firm Age	19.42	24.00	12.46	11.00	<0.001
Market-to-Book	2.85	2.04	2.92	1.94	0.013

(continued on next page)

Panel B: Distribution by Firm Characteristics

	Sample (n = 214)		2003 Compustat (excluding sample firms)[b]		
	Mean	**Median**	**Mean**	**Median**	**Median Comparison[c]**
ROA	0.036	0.032	−0.229	0.012	<0.001
Segments	3.53	3.00	2.86	2.00	<0.001

Panel C: Distribution of Sample Firms with Material Weaknesses by Industry

	Sample		Doyle et al. (2007a) Sample	
Industry[d]	**n**	**%**	**n**	**%**
Banks and Insurance	11	24.44	147	13.3
Durable Manufacturers	6	13.33	198	18.0
Utilities	6	13.33	35	3.2
Computers	5	11.11	208	18.9
Retail	4	8.89	137	12.4
Textiles, Printing, and Publishing	5	11.11	41	3.7
Transportation	3	6.67	64	5.8
Services	3	6.67	106	9.6
Drug and Medical Equipment	1	2.22	72	6.5
Food	1	2.22	16	1.4
Refining and Extractive	—	—	32	2.9
Mining and Construction	—	—	21	1.9
Chemicals	—	—	21	1.9
Miscellaneous	—	—	4	0.4
Total	45	100.00	1,102	100.0

[a] Industry affiliations are determined using the following SIC codes: Agriculture: 100–999; Mining and Construction: 1000–1299, 1400–1999; Food: 2000–2111; Textiles, Printing, and Publishing: 2200–2799; Chemicals: 2800–2824, 2840–2899; Drug and Medical Equipment: 2830–2836, 3840–3851; Refining and Extractive: 1300–1399, 2900–2999; Durable Manufacturers: 3000–3569, 3580–3669, 3680–3839, 3852–3999; Computers: 3570–3579, 3670–3679, 7370–7379; Transportation: 4000–4899; Utilities: 4900–4999; Retail: 5000–5999; Banks and Insurance: 6000–6999; Services: 7000–7369, 7380–8999; and Miscellaneous: 9000–9999. The benchmark group consists of all publicly traded firms on 2003 Compustat with non-missing market value of equity (#199 times #25), total assets (#6), and earnings before extraordinary items (#18).

[b] The benchmark group consists of a maximum of 6,803 firm observations from 2003 Compustat database after excluding our sample firms. Market Value is the market value of equity (#199 times #25); Total Assets (#6); Firm Age is the number of years the firm has Compustat data since 1980 and up to 2003; Market-to-Book is the market value of equity

(continued on next page)

(#199 times #25) divided by common equity (#60), winsorized at top and bottom 1 percent; ROA is earnings before extraordinary items (#18) divided by average total assets (#6), Winsorized at top and bottom 1 percent; Segments is the sum of the number of operating and geographic segments reported by the Compustat Segments database.

[c] Reported p-values are from two-tailed Wilcoxon rank sum tests.

[d] Industry affiliations are determined using the following SIC codes: Mining and Construction: 1000–1299, 1400–1999; Food: 2000–2111; Textiles, Printing, and Publishing: 2200–2799; Chemicals: 2800–2824, 2840–2899; Drug and Medical Equipment: 2830–2836, 3840–3851; Refining and Extractive: 1300–1399, 2900–2999; Durable Manufacturers: 3000–3569, 3580–3669, 3680–3839, 3852–3999; Computers: 3570–3579, 3670–3679, 7370–7379; Transportation: 4000–4899; Utilities: 4900–4999; Retail: 5000–5999; Banks and Insurance: 6000–6999; Services: 7000–7369, 7380–8999; and Miscellaneous (9000–9999). There are no material weakness firm observations in Agriculture (100–999) for both our sample and Doyle et al.'s (2007a) sample.

TABLE 4

Descriptive Statistics for IAF Attributes, IAF Activities, and Control Variables, Partitioned by Whether Firm Disclosed a Material Weakness[a]

	Full Sample (n = 214)						
	Mean	**Std. Dev.**	**Min**	**Q1**	**Median**	**Q3**	**Max**
EXPERIENCE	8.901	4.343	1.000	6.000	8.000	11.000	24.000
EDUCATION	4.467	0.503	2.000	4.200	4.509	4.750	6.000
CERTIFICATIONS	0.625	0.236	0.000	0.500	0.613	0.800	1.000
TRAINING	53.640	28.600	0.000	40.000	48.000	74.000	178.000
CAEAC	0.822	0.181	0.000	0.750	0.875	1.000	1.000
CAEOFFICER	0.388	0.488	0.000	0.000	0.000	1.000	1.000
IASIZE	0.031	0.028	0.001	0.012	0.022	0.038	0.160
FIELDWORKQA	1.318	0.360	0.000	1.143	1.357	1.571	2.000
IAGRADE	0.467	0.500	0.000	0.000	0.000	1.000	1.000
FINANCIALFOCUS	1.185	0.447	0.000	0.833	1.167	1.500	2.000
FOLLOWUP	0.804	0.398	0.000	1.000	1.000	1.000	1.000
COORDINATION	0.883	0.322	0.000	1.000	1.000	1.000	1.000
SEGMENTS	3.528	2.764	1.000	1.000	3.000	5.000	19.000
FOREIGNTRANSACTIONS	0.299	0.459	0.000	0.000	0.000	1.000	1.000
M&A	0.193	0.243	0.000	0.000	0.200	0.400	1.000
RESTRUCTURE	0.379	0.367	0.000	0.000	0.400	0.600	1.000
SALESGROWTH	0.105	0.166	0.000	0.000	0.000	0.200	1.000
INVENTORY	0.075	0.095	0.000	0.009	0.041	0.108	0.602
MARKETVALUE ($M)	14,542	31,946	90	1,763	4,683	11,610	276,168
LOSS	0.121	0.231	0.000	0.000	0.000	0.200	1.000
CFO	0.088	0.057	−0.100	0.050	0.081	0.122	0.254
SHUMWAY	0.004	0.010	0.000	0.001	0.002	0.004	0.139
AGE	19.421	6.440	4.000	14.000	24.000	24.000	24.000
AUDITORSPECIALIST	0.383	0.487	0.000	0.000	0.000	1.000	1.000
BLUERIBBONAC	0.973	0.093	0.100	1.000	1.000	1.000	1.000
ISO9000	0.416	0.494	0.000	0.000	0.000	1.000	1.000

(continued on next page)

	Full Sample (n = 214)						
	Mean	Std. Dev.	Min	Q1	Median	Q3	Max
RESTATEMENT	0.706	0.457	0.000	0.000	1.000	1.000	1.000
AUDITORRESIGN	0.042	0.201	0.000	0.000	0.000	0.000	1.000
INSTITUTIONALOWNERSHIP	0.461	0.135	0.098	0.370	0.465	0.547	0.917
REGULATEDINDUSTRY	0.355	0.480	0.000	0.000	0.000	1.000	1.000

	Material Weakness Firms (n = 45)		Non-MW Firms (n = 169)		p-values for Mean/Median Comparisons[b]	
	Mean	Median	Mean	Median	Mean	Median
EXPERIENCE	9.420	9.000	8.763	8.000	0.369	0.285
EDUCATION	4.379	4.400	4.491	4.542	0.187	0.226
CERTIFICATIONS	0.652	0.615	0.618	0.600	0.397	0.436
TRAINING	52.311	45.000	53.994	50.000	0.767	0.409
CAEAC	0.786	0.875	0.831	0.875	0.136	0.193
CAEOFFICER	0.422	0.000	0.379	0.000	0.596	0.596
IASIZE	0.030	0.021	0.032	0.022	0.693	0.944
FIELDWORKQA	1.244	1.143	1.338	1.429	0.196	0.145
IAGRADE	0.622	1.000	0.426	0.000	0.019	0.019
FINANCIALFOCUS	1.067	1.000	1.216	1.167	0.046	0.063
FOLLOWUP	0.756	1.000	0.817	1.000	0.362	0.362
COORDINATION	0.933	1.000	0.870	1.000	0.168	0.241
SEGMENTS	3.222	2.000	3.609	3.000	0.405	0.228
FOREIGNTRANSACTIONS	0.400	0.000	0.272	0.000	0.097	0.097
M&A	0.244	0.200	0.179	0.200	0.108	0.229
RESTRUCTURE	0.476	0.400	0.354	0.200	0.048	0.082
SALESGROWTH	0.111	0.000	0.103	0.000	0.776	0.558
INVENTORY	0.076	0.022	0.075	0.043	0.932	0.330
MARKETVALUE ($M)	7,158	2,863	16,508	5,089	0.005	0.021
LOSS	0.227	0.000	0.093	0.000	0.008	0.003
CFO	0.060	0.056	0.095	0.085	<0.0001	0.000

(continued on next page)

	Material Weakness Firms (n = 45)		Non-MW Firms (n = 169)		p-values for Mean/Median Comparisons[b]	
	Mean	Median	Mean	Median	Mean	Median
SHUMWAY	0.008	0.003	0.003	0.002	0.082	0.057
AGE	19.444	24.000	19.414	24.000	0.978	0.935
AUDITORSPECIALIST	0.356	0.000	0.391	0.000	0.670	0.670
BLUERIBBONAC	0.971	1.000	0.973	1.000	0.885	0.704
ISO9000	0.289	0.000	0.450	0.000	0.052	0.053
RESTATEMENT	0.667	1.000	0.716	1.000	0.521	0.521
AUDITORRESIGN	0.111	0.000	0.024	0.000	0.079	0.010
INSTITUTIONALOWNERSHIP	0.456	0.470	0.463	0.460	0.749	0.955
REGULATEDINDUSTRY	0.378	0.000	0.349	0.000	0.723	0.723

[a] See Table 2 for variable definitions.
[b] Reported p-values for means are from two-tailed t-tests; reported p-values for medians are from two-tailed Wilcoxon rank sum tests.

TABLE 5

Spearman/Pearson Correlations Among Measures of IAF Quality and Control Variables (n = 214)

Panel A: Spearman/Pearson Correlations among Measures of IAF Quality

	MW	*EXPERIENCE*	*EDUCATION*	*CERTIFICATION*	*TRAINING*	*CAEAC*
MW		0.06	−0.09	0.06	−0.02	−0.10
EXPERIENCE	0.07		−0.11	**0.38**	−0.09	0.01
EDUCATION	−0.08	−0.10		−0.02	**0.15**	−0.04
CERTIFICATIONS	0.05	**0.39**	−0.01		**−0.14**	0.02
TRAINING	−0.06	−0.08	**0.22**	−0.10		*0.13*
CAEAC	−0.09	0.04	0.00	0.04	*0.12*	
CAEOFFICER	0.04	−0.03	−0.06	−0.07	*0.13*	*0.13*
IASIZE	−0.01	−0.10	*0.13*	−0.08	**−0.14**	*0.12*
FIELDWORKQA	−0.10	**−0.18**	0.00	−0.05	**0.18**	**0.19**
IAGRADE	**0.16**	−0.01	0.01	−0.11	0.06	*0.13*
FINANCIALFOCUS	*−0.13*	0.04	0.02	0.01	**0.16**	**0.35**
FOLLOWUP	−0.06	0.01	−0.01	0.04	0.04	*0.12*
COORDINATION	0.08	−0.10	−0.01	−0.03	**0.17**	**0.14**

	CAEOFFICER	*IASIZE*	*FIELDWORKQA*	*IAGRADE*	*FINANCIALFOCUS*	*FOLLOWUP*	*COORDINATION*
MW	0.04	−0.03	−0.11	**0.16**	**−0.14**	−0.06	0.08
EXPERIENCE	−0.06	0.05	**−0.26**	−0.02	0.01	0.01	*−0.13*
EDUCATION	−0.06	**0.15**	−0.01	−0.03	0.00	−0.03	−0.02
CERTIFICATIONS	−0.07	0.00	−0.10	−0.09	0.00	0.05	−0.04
TRAINING	0.09	**−0.14**	**0.15**	0.06	**0.16**	0.04	**0.17**
CAEAC	*0.12*	*0.12*	**0.22**	*0.12*	**0.34**	*0.13*	**0.20**
CAEOFFICER		**−0.21**	*0.13*	**0.18**	−0.04	0.08	0.08
IASIZE	**−0.22**		−0.11	*−0.12*	0.01	0.01	−0.10
FIELDWORKQA	−0.10	−0.10		**0.20**	**0.26**	*0.13*	**0.23**
IAGRADE	−0.04	0.06	**0.19**		*0.12*	0.09	0.02
FINANCIALFOCUS	0.08	*0.12*	**0.23**	0.11		0.02	*0.13*
FOLLOWUP	0.06	0.05	**0.15**	0.09	0.02		**0.15**
COORDINATION	−0.11	−0.10	**0.19**	0.02	*0.12*	0.15	

(continued on next page)

Panel B: Spearman Correlations among Measures of IAF Quality and Control Variables

	MW	*EXPERIENCE*	*EDUCATION*	*CERTIFICATION*	*TRAINING*	*CAEAC*
SEGMENTS	−0.08	−0.07	**0.15**	−0.03	−0.07	0.02
FOREIGNTRANSACTIONS	*0.11*	**−0.20**	0.11	−0.11	*−0.13*	−0.02
M&A	0.08	−0.03	0.02	−0.03	0.06	0.06
RESTRUCTURE	*0.12*	**−0.18**	0.07	*−0.13*	0.02	0.08
SALESGROWTH	0.04	**0.16**	−0.02	0.06	0.04	0.02
INVENTORY	−0.07	**−0.14**	0.01	*−0.11*	0.03	−0.05
MARKETVALUE	**−0.16**	**−0.22**	**0.17**	**−0.19**	**0.31**	**0.15**
LOSS	**0.20**	−0.11	0.04	0.02	−0.04	−0.06
CFO	**−0.26**	*−0.13*	*0.12*	−0.01	−0.10	0.04
SHUMWAY	*0.13*	0.04	0.00	−0.05	**0.17**	0.00
AGE	−0.01	−0.05	0.04	−0.09	**0.16**	−0.09
AUDITORSPECIAL	−0.03	0.02	0.05	−0.01	*0.13*	0.08
BLUERIBBONAC	−0.03	−0.03	−0.03	−0.07	**0.17**	**0.17**
ISO9000	*−0.13*	**−0.27**	0.09	**−0.18**	0.11	0.02
RESTATEMENT	−0.04	−0.02	0.03	0.04	0.08	0.07
AUDITORRESIGN	**0.18**	0.02	−0.06	0.04	−0.04	0.08
INSTITUTIONOWNER	0.00	0.06	0.01	0.03	**−0.14**	−0.03
REGULATEDINDUSTRY	0.02	**0.28**	−0.09	0.11	**0.14**	0.02

	CAEOFFICER	*IASIZE*	*FIELDWORKQA*	*IAGRADE*	*FINANCIALFOCUS*	*FOLLOWUP*	*COORDINATION*
SEGMENTS	**0.19**	−0.01	0.01	0.02	0.01	0.00	0.02
FOREIGNTRANSACTIONS	**0.20**	−0.02	0.08	0.10	−0.03	−0.04	−0.02
M&A	**0.14**	−0.03	−0.07	0.08	−0.02	0.00	−0.04
RESTRUCTURE	**0.18**	−0.04	0.06	**0.16**	−0.10	0.01	0.05
SALESGROWTH	0.02	0.07	**−0.15**	0.02	*0.12*	−0.02	−0.10
INVENTORY	**0.42**	0.10	−0.07	0.05	−0.09	0.06	−0.07
MARKETVALUE	**−0.28**	**−0.46**	**0.26**	**0.16**	0.05	0.06	**0.18**
LOSS	−0.01	−0.10	0.09	0.11	−0.03	−0.02	−0.02
CFO	**0.40**	0.09	−0.05	**−0.18**	0.02	−0.06	−0.10
SHUMWAY	**−0.20**	−0.05	**0.15**	**0.16**	0.06	0.06	**0.15**
AGE	0.03	**−0.15**	**0.14**	−0.02	−0.07	0.05	0.02
AUDITORSPECIAL.	−0.09	−0.09	0.10	0.01	−0.04	**0.17**	**0.14**

(continued on next page)

	CAEOFFICER	*IASIZE*	*FIELDWORKQA*	*IAGRADE*	*FINANCIALFOCUS*	*FOLLOWUP*	*COORDINATION*
BLUERIBBONAC	−0.08	−0.07	**0.24**	0.04	0.10	−0.02	**0.20**
ISO9000	**0.29**	0.01	0.11	0.07	−0.02	0.04	−0.05
RESTATEMENT	**−0.18**	*−0.13*	0.08	0.01	−0.05	0.04	−0.04
AUDITORRESIGN	0.06	0.09	0.02	0.04	0.07	0.05	0.00
INSTITUTIONOWNER	**0.20**	0.05	−0.05	*−0.13*	−0.09	0.01	0.05
REGULATEDINDUSTRY	**−0.50**	−0.02	−0.01	0.03	0.03	0.05	0.06

Panel C: Pearson Correlations among Measures of IAF Quality and Control Variables

	MW	*EXPERIENCE*	*EDUCATION*	*CERTIFICATION*	*TRAINING*	*CAEAC*
SEGMENTS	−0.08	−0.09	*0.12*	−0.05	−0.08	0.01
FOREIGNTRANSACTIONS	*0.11*	**−0.16**	*0.12*	−0.09	−0.09	−0.04
M&A	0.11	−0.05	0.04	−0.05	0.09	0.07
RESTRUCTURE	**0.14**	**−0.16**	0.07	−0.11	0.02	0.04
SALESGROWTH	0.02	**0.14**	0.01	0.05	0.02	0.02
INVENTORY	0.01	−0.05	−0.06	−0.03	−0.03	0.00
MARKETVALUE	**−0.17**	**−0.21**	**0.16**	**−0.17**	**0.26**	**0.18**
LOSS	**0.24**	−0.08	0.00	0.09	−0.06	−0.04
CFO	**−0.25**	−0.08	0.11	−0.02	*−0.13*	0.04
SHUMWAY	**0.23**	−0.03	−0.01	0.01	−0.10	−0.02
AGE	0.01	−0.05	−0.02	−0.07	*0.12*	−0.11
AUDITORSPECIAL.	−0.03	0.01	0.06	−0.01	0.10	0.08
BLUERIBBONAC	−0.01	−0.04	−0.03	−0.01	0.10	**0.31**
ISO9000	*−0.13*	**−0.20**	0.04	**−0.18**	0.11	0.00
RESTATEMENT	−0.04	−0.06	0.01	0.06	0.11	0.07
AUDITORRESIGN	**0.18**	0.01	0.01	0.04	−0.04	0.08
INSTITUTIONOWNER	−0.02	0.06	0.08	0.07	−0.07	−0.02
REGULATEDINDUSTRY	0.02	**0.19**	−0.10	0.11	0.11	0.04

	CAEOFFICER	*IASIZE*	*FIELDWORKQA*	*IAGRADE*	*FINANCIALFOCUS*	*FOLLOWUP*	*COORDINATION*
SEGMENTS	−0.09	0.10	0.01	0.03	0.01	0.00	0.02
FOREIGNTRANSACTIONS	*−0.12*	0.11	0.07	0.10	−0.03	−0.04	−0.02
M&A	−0.05	0.03	0.00	0.06	0.01	−0.01	−0.05

(continued on next page)

	CAEOFFICER	*IASIZE*	*FIELDWORKQA*	*IAGRADE*	*FINANCIALFOCUS*	*FOLLOWUP*	*COORDINATION*
RESTRUCTURE	−0.01	0.06	0.04	**0.19**	−0.10	0.00	0.04
SALESGROWTH	0.03	−0.06	−0.11	0.02	0.09	0.01	−0.05
INVENTORY	−0.10	**0.35**	**−0.14**	−0.02	−0.11	−0.02	−0.07
MARKETVALUE	**0.23**	**−0.35**	**0.25**	**0.15**	0.05	0.07	**0.17**
LOSS	**−0.14**	0.01	0.07	**0.14**	−0.03	−0.03	−0.09
CFO	**−0.20**	**0.36**	−0.04	**−0.18**	0.04	−0.03	−0.11
SHUMWAY	0.01	−0.09	0.04	*0.13*	−0.10	−0.09	0.00
AGE	*−0.13*	**−0.14**	*0.12*	−0.06	−0.08	0.03	0.05
AUDITORSPECIAL.	0.06	**−0.14**	0.09	0.01	−0.04	**0.17**	**0.14**
BLUERIBBONAC	0.02	−0.03	**0.25**	0.10	0.03	0.11	**0.22**
ISO9000	−0.09	**0.14**	0.09	0.07	0.00	0.04	−0.05
RESTATEMENT	0.05	**−0.20**	0.06	0.01	−0.03	0.04	−0.04
AUDITORRESIGN	−0.07	0.09	0.00	0.04	0.08	0.05	0.00
INSTITUTIONOWNER	*−0.13*	**0.15**	−0.07	*−0.12*	*−0.12*	−0.01	0.05
REGULATEDINDUSTRY	**0.21**	**−0.41**	0.03	0.03	0.03	0.05	0.06

Bold = p-value < 0.05. Italic = p-value < 0.10.
See Table 2 for variable definitions.

suggesting that IAFs that focus greater attention on financial reporting activities tend to report more information to the audit committee. All other significant correlations among our IAF quality measures fall below 0.30. One measure of objectivity, *CAEOFFICER*, is significantly associated with three control variables, *INVENTORY* (Spearman r = 0.42), *CFO* (Spearman r = 0.40), and *REGULATEDINDUSTRY* (Spearman r = −0.50). Investment in the IAF (*IASIZE)* is significantly and negatively correlated with the two control variables, *MARKETVALUE* (Spearman r = −0.46; Pearson r = −0.35) and *REGULATEDINDUSTRY* (Pearson r = −0.41). Remaining correlations between IAF quality measures and control variables are below 0.40.

Multivariate Results

Table 6 reports the results of our logistic regression. We report parameter estimates, marginal effects, and the change in the probability of a firm disclosing a material weakness as a result of moving from the first to the third quartile value of the independent variable, while holding other regressors at their mean values.[19] We report the results for full and stepwise regression models in Panels A and B, respectively. All p-values refer to two-tailed tests of significance.

We find little support for H1, which predicts associations between MW disclosures and the IAF quality attributes of competence, objectivity, and IAF investment. Among our four proxies for IAF competence, only *EDUCATION* is significantly associated with the probability that a firm reports a material weakness (Coeff. = −1.682, $p < 0.05$). The results in column (4) indicate that moving from first to third quartile of *EDUCATION* decreases the probability of material weakness disclosure by 1.8 percent. Contrary to H1, the remaining measures of IAF competence (*EXPERIENCE*, *CERTIFICATION*, and *TRAINING*) and our proxies for IAF objectivity (*CAEAC* or *CAEOFFICER*) and IAF investment (*IASIZE*) are not significantly associated with MW disclosures. The lack of a statistically significant association could reflect three factors. First, IAF competence, objectivity, and investment improve the IAF's ability to prevent material weaknesses from occurring, while also increasing the likelihood that existing material weaknesses are detected and disclosed. Hence, these opposing effects may offset each other. Second, the small size and relative homogeneity of our sample, combined with the large number of control variables, lowers the power of our statistical tests. Finally, *CAEAC* and *IASIZE* are significantly correlated with multiple control variables that are significant predictors of MWs, making it difficult to isolate relations between these IAF attributes and MW disclosures. Accordingly, further investigation into these specific aspects of IAF quality is warranted.

We find strong support for H2, which predicts associations between MW disclosures and the following IAF activities: *FIELDWORKQA*, *FINANCIALFOCUS*, and *IAGRADE*. The coefficients on *FIELDWORKQA* and *FINANCIALFOCUS* are negative and significant (Coeff. = −2.674, $p < 0.01$; Coeff. = −1.955, $p < 0.05$, respectively), indicating that MW disclosures are decreasing in the use of quality assurance techniques during fieldwork and the extent to which IAF scope includes financial reporting processes. Moving from first to third quartile of *FIELDWORKQA* and *FINANCIALFOCUS* decreases the probability of firms reporting a material weakness by 2.1 percent and 2.7 percent, respectively. We also find that *IAGRADE* is positively and significantly associated with the probability of disclosing a material weakness (Coeff. = 1.445, $p < 0.05$), suggesting that grades facilitate Section 404 risk assessment, and thereby increase the likelihood that extant MWs are identified. Firms whose internal auditors issue grades or ratings are 3.1 percent more likely to report a material weakness than those that do not.

[19] We compute the marginal effects as $e^{\beta' X} / (1 + e^{\beta' X})^2$ where $\beta' X$ is computed using the mean value of X. If X contains logs, then we substitute the mean of the original variable into the logarithm function.

TABLE 6

Logistic Model Estimating the Relation between IAF Attributes and Activities and the Likelihood of a Material Weakness Disclosure[a]

Panel A: Full Model

	Coeff. (χ^2) (1)	Marginal Effect[c] (2)	Change in Prob. Q1 versus Q3 Values[d] (3)
Intercept	6.780		
	(2.36)		
IAF Attributes			
EXPERIENCE	0.001	0.000	0.000
	(0.00)		
EDUCATION	−1.682**	−0.032	−0.018
	(6.19)		
CERTIFICATIONS	0.933	0.018	0.005
	(0.53)		
TRAINING	0.001	0.000	0.001
	(0.01)		
CAEAC	−1.660	−0.031	−0.007
	(0.97)		
CAEOFFICER	0.412	0.008	0.008
	(0.47)		
IASIZE	6.523	0.124	0.003
	(0.18)		
IAF Activities			
FIELDWORKQA	−2.674***	−0.051	−0.021
	(7.93)		
IAGRADE	1.445**	0.027	0.031
	(5.64)		
FINANCIALFOCUS	−1.955**	−0.037	−0.027
	(6.12)		
FOLLOWUP	−1.213*	−0.023	−0.034

(continued on next page)

Panel A: Full Model

	Coeff. (χ^2) (1)	Marginal Effect[c] (2)	Change in Prob. Q1 versus Q3 Values[d] (3)
	(3.63)		
COORDINATION	4.095***	0.078	0.030
	(8.00)		
Control Variables			
SEGMENTS	−1.465***	−0.028	−0.099
	(8.95)		
FOREIGNTRANSACTIONS	3.650***	0.069	0.196
	(15.13)		
M&A	4.629***	0.088	0.041
	(11.60)		
RESTRUCTURE	1.137	0.022	0.012
	(1.37)		
SALESGROWTH	−1.093	−0.021	−0.004
	(0.42)		
INVENTORY	−7.645**	−0.145	−0.017
	(3.88)		
MARKETVALUE	−0.387	−0.007	−0.022
	(2.15)		
LOSS	1.793	0.034	0.007
	(1.33)		
CFO	−28.061***	−0.532	−0.046
	(9.87)		
SHUMWAY	210.300**	3.987	0.008
	(4.11)		
AGE	2.327***	0.044	0.022
	(7.43)		
AUDITORSPECIALIST	0.863	0.016	0.019
	(1.89)		
BLUERIBBONAC	0.529	0.010	0.000

(continued on next page)

Panel A: Full Model

	Coeff. (χ^2) (1)	Marginal Effect[c] (2)	Change in Prob. Q1 versus Q3 Values[d] (3)
	(0.04)		
ISO9000	−4.006***	−0.076	−0.093
	(14.84)		
RESTATEMENT	−0.401	−0.008	−0.008
	(0.45)		
AUDITORRESIGN	5.492***	0.104	0.776
	(12.95)		
INSTITUTIONALOWNERSHIP	−1.022	−0.019	−0.003
	(0.19)		
REGULATEDINDUSTRY	−1.723*	−0.033	−0.029
	(2.98)		
Likelihood Ratio χ^2	107.15		
Model Significance	<0.0001		
Max-Rescaled R^2 (%)	61.31		
Number of *MW* Obs.	45		
Total Number of Obs.	214		

Panel B: Stepwise Model[b]

	Coeff. (χ^2) (4)	Marginal Effect[c] (5)	Change in Prob. Q1 versus Q3 Values[d] (6)
Intercept	7.720		
	(5.09)		
IAF Attributes			
EDUCATION	−1.579***	−0.053	−0.029
	(7.19)		

(continued on next page)

Panel B: Stepwise Model[b]

	Coeff. (χ^2) (4)	Marginal Effect[c] (5)	Change in Prob. Q1 versus Q3 Values[d] (6)
IAF Activities			
FIELDWORKQA	−2.324***	−0.078	−0.032
	(9.02)		
IAGRADE	1.500***	0.050	0.056
	(6.81)		
FINANCIALFOCUS	−2.107***	−0.070	−0.052
	(9.94)		
FOLLOWUP	−1.137*	−0.038	−0.054
	(3.67)		
COORDINATION	3.163***	0.106	0.047
	(7.72)		
Control Variables			
SEGMENTS	−1.376***	−0.046	−0.147
	(9.66)		
FOREIGNTRANSACTIONS	3.642***	0.122	0.303
	(16.30)		
M&A	4.311***	0.144	0.065
	(12.03)		
RESTRUCTURE	1.196	0.040	0.022
	(1.90)		
INVENTORY	−7.720**	−0.258	−0.029
	(5.33)		
MARKETVALUE	−0.487**	−0.016	−0.031
	(4.92)		
CFO	−29.530***	−0.986	−0.085
	(13.65)		
SHUMWAY	238.700**	7.974	0.016
	(7.32)		

(continued on next page)

Panel B: Stepwise Model[b]

	Coeff. (χ^2) (4)	Marginal Effect[c] (5)	Change in Prob. Q1 versus Q3 Values[d] (6)
AGE	2.226***	0.074	0.037
	(8.18)		
AUDITORSPECIALIST	0.721	0.024	0.026
	(1.71)		
ISO9000	−3.928***	−0.131	−0.152
	(16.90)		
AUDITORRESIGN	5.102***	0.170	0.798
	(13.73)		
REGULATEDINDUSTRY	−2.183**	−0.073	−0.064
	(6.39)		
Likelihood Ratio χ^2	102.88		
Model Significance	<0.0001		
Max-Rescaled χ^2 (%)	59.41		
Number of *MW* Obs.	45		
Total Number of Obs.	214		

*, **, *** $\Pr > \chi^2$ of 0.1, 0.05, and 0.01, respectively.

[a] See Table 2 for variable definitions.

[b] The significance level of $\chi 2$ for a variable to stay in the model is 0.30.

[c] The marginal effects are computed as $e^{\beta' X} / (1 + e^{\beta' X})2$, where $\beta' X$ is computed using the mean values of X. If X contains logs, then we substitute means of original variables into log functions.

[d] Tabled values in the Change in Probability column show the change in the probability of a firm disclosing a material weakness as a result of moving from the first to the third quartile value of the variable of interest or from 0 to 1 if the variable is an indicator variable, holding all other variables constant at their mean values.

As predicted by H3, firms with IAFs that follow-up on extant control problems are significantly (Coeff. = -1.213, $p < 0.10$) less likely to report a material weakness. This result suggests that monitoring the remediation of previously identified control problems prompts management to correct them in a timely manner. A one-unit change in *FOLLOWUP* is associated with a 3.4 percent decrease in the probability of reporting a material weakness.

Consistent with H4, firms are significantly more likely to report a material weakness (Coeff. = 4.095, $p < 0.01$) when the IAF coordinates audit activities with the external auditors. This result suggests that external auditors can increase the effectiveness of their Section 404 procedures by coordinating with a firm's IAF. The change in probability analysis shows that a unit change in *COORDINATION* increases the probability of firms reporting a material weakness by 3 percent.

Our findings regarding the nature and scope of IAF activities suggest implications for practice. The results for *FIELDWORKQA* and *FINANCIALFOCUS* indicate that these audit practices prevent MWs from occurring and/or ensure that extant MWs are detected early enough in the firm's Section 404 testing to enable management to correct problems. The year-end timing of most Section 404 work suggests that these activities are more preventative than detective. Together with tests of the *FOLLOWUP* variable, our results imply that firms can potentially improve ICFR by expanding fieldwork quality assurance practices, devoting greater effort to financial statement reporting processes, and aggressively following up on previously detected audit exceptions.

The positive relations between MW disclosures and both *IAGRADE* and *COORDINATION* suggest that these activities increase the effectiveness of Section 404 compliance processes by facilitating risk assessment. IAF grading promotes rapid assessment and prioritization of control risks by management, audit committees, and external auditors (PwC 2006). Similarly, information gained through coordination enables external auditors to alter the nature, timing, and extent of their testing to reflect risk levels. Accordingly, these findings confirm the appropriateness of recent regulatory guidance that instructs firms and external auditors to use a risk-based approach to Section 404 compliance (SEC 2007; PCAOB 2005, 2007b). The results are also consistent with recent PCAOB (2005, 2007a) guidance that encourages external auditors to more effectively use the work of others.[20]

The signs and significance levels of our control variables are generally consistent with prior research, except for the negative coefficient on *SEGMENTS* and the positive coefficient on *AGE* (Ashbaugh-Skaife et al. 2007; Doyle et al. 2007a; Krishnan and Vishwanathan 2007; Stephens 2009). We attribute these differences to our sample, which has a much larger proportion of utilities and financial firms than samples used in prior studies.

Since many independent variables are not statistically significant in Panel A of Table 6 and we have a relative small sample size, we also present logit results for a stepwise hierarchical regression model in Panel B.[21] The coefficient signs in the reduced model are consistent with the full model, while the sizes, significance levels, and marginal effects of the coefficients are generally larger.

Sensitivity Analysis

One limitation of our study is that the IAF data describe different years for different firms. IAF data for 57 firms in the sample correspond to the pre-SOX time period. This could bias our

---

[20] The methods of coordination used (multiple methods per firm are permitted) and the percentage of firms using each method are as follows: (1) loan to external auditors, 28 percent, (2) perform complete or partial audit of specific locations, products or functions, 58 percent, (3) conduct joint annual planning sessions, 58 percent, and (4) conduct joint risk or control sessions, 39 percent. We classify IAFs that responded "yes" to (1) and/or (2) as performing work for use by external auditors. We perform robustness tests with a continuous measure based on the number of coordination methods used. This variable gives similar results (Coeff.= 2.179, $p = 0.046$). However, multiple methods of coordination are not necessarily indicative of more coordination; thus, we use an indicator variable in our main results.

[21] We used a p-value of 0.30 as a discretionary cutoff for a variable to be included in the model. Additional sensitivity tests using cut-off p-values of 0.2, 0.4, and 0.5 produce materially similar results.

results if the year of IAF data collection is correlated with both MW disclosures and our measures of IAF quality. To test for bias, we re-estimate Equation (1) after including an indicator variable, *SOX*, to indicate whether IAF data were collected on or after July 2002. The coefficient on *SOX* (Coeff. = −0.418) is not statistically significant (p = 0.509), and inclusion of the *SOX* variable does not materially affect our results. For firms with survey data available for both 2003 and 2004, we use the data from 2003 (rather than 2004) and re-estimate the models. The results are largely consistent with our reported results.

We further confirm the robustness of our results by using additional control variables for external monitoring. We include an additional corporate governance factor, Gompers et al.'s (2003) *GINDEX*. *GINDEX* is measured as the average G-index score for 2002 and 2004. This reduces the sample size to 200 firms. We obtain similar results for all IAF variables except *IAGRADE* and *FOLLOWUP*. In the full model, the coefficients for *IAGRADE* and *FOLLOWUP* are insignificant. We also control for the possibility that a firm audited by one of the Big 4 audit firms may be more likely to disclose a material weakness. Our results remain unchanged from those reported in Table 6.

VI. CONCLUSION

This study investigates associations between material weakness disclosures and various IAF attributes and activities using survey data collected by the IIA. Our results indicate that the nature and scope of IAF activities are more strongly associated with MW disclosures than the IAF attributes of competence, objectivity, and investment. Among IAF attribute measures, only the education level of the IAF is significantly associated with MW disclosures. Regarding IAF activities, we find that MW disclosures are negatively associated with the extent to which the IAF uses QA techniques in fieldwork, audits activities related to financial reporting, and follows up on previously identified control problems. The year-end timing of most Section 404 work and the nature of follow-up procedures suggests that these activities are more likely to be preventative rather than detective. We also find that MW disclosures are positively related with both IAF grading of audit engagements and external-internal auditor coordination. We interpret this finding to indicate that these activities increase the effectiveness of Section 404 compliance processes by facilitating risk assessment, consistent with the risk-based approach promoted by regulatory guidance (SEC 2007; PCAOB 2005, 2007a). Together, our results have important implications for managers who determine IAF staffing and activities, standard-setters who provide auditing guidance, and external auditors responsible for Section 404 work.

This study makes several important contributions to the literature. First, we expand extant research on both internal auditing and ICFR by using survey data from companies to document associations between various measures of IAF quality and material weakness disclosures. Research on the determinants of internal control deficiencies has investigated firm characteristics such as size, profitability, and complexity, without consideration of the role of the IAF (Ashbaugh-Skaife et al. 2007; Doyle et al. 2007a). With the exception of Prawitt et al. (2009), prior IAF studies focus on how measures of IAF quality affect the quality of internal auditors' decisions, without linking IAF quality to an actual outcome measure of control effectiveness, such as a material weakness (Berry et al. 1987; Harrell et al. 1989; Church and Schneider 1995). Accordingly, this study complements Prawitt et al. (2009), which finds a positive relation between a comprehensive measure of IAF quality and earnings quality. In combination, these studies suggest that the IAF is an important component of the financial reporting process. Second, the bi-directional relations between IAF quality measures and MW disclosures provide evidence on the roles that specific IAF attributes and activities play in the existence and detection of MWs. Few archival studies directly link auditor practices to the prevention or detection of audit exceptions. Finally, we provide evidence that external auditors are more likely to detect material weaknesses

when they coordinate their efforts with the IAF. While a significant branch of the internal auditing literature focuses on the relation between external auditors and the IAF, no study of which we are aware has examined whether IAF involvement in external audits increases the effectiveness of external audits.

Our study is subject to several limitations. Most notably, the small size and homogeneity of our sample, combined with the large number of control variables, lowers the power of our statistical tests. Accordingly, we cannot determine whether the lack of statistically significant hypothesized relations between IAF attributes is due to low statistical power or competing effects on the existence and detection of control problems. The stepwise tests mitigate this problem but do not eliminate it. Furthermore, large firms with relatively sophisticated IAFs tend to participate in the GAIN survey. This limits our ability to generalize findings to firms that did not respond to the GAIN survey. Despite these limitations, this study increases our understanding of the IAF's role in ICFR.

REFERENCES

Albrecht, W. S., and C. O. Albrecht. 2004. *Fraud Examination and Prevention*. Mason, Ohio: Thompson/South-Western.

American Institute of Certified Public Accountants (AICPA). 1991. *The Auditor's Consideration of the Internal Audit Function in an Audit of Financial Statements*. Statement on Auditing Standards No. 65. New York, NY: AICPA.

Ashbaugh-Skaife, H., D. Collins, and W. Kinney. 2007. The discovery and reporting of internal control deficiencies prior to SOX-mandated audits. *Journal of Accounting and Economics* 44 (1–2): 166–192.

Asare, S. K., L. A. Cunningham, and A. Wright. 2007. The Sarbanes-Oxley Act: Legal implications and research opportunities. Research Paper 122, Boston College Law School Legal Studies Research Paper Series. Boston, MA: Boston College.

Bailey, A. D., Jr., A. A. Gramling, and S. Ramamoorti, eds. 2003. *Research Opportunities in Internal Auditing*. Altamonte Springs, FL: IIA Research Foundation.

Bedard, J. C., and L. Graham. 2011. Factors affecting the severity of Sarbanes-Oxley Section 404 internal control deficiencies: Archival evidence. *The Accounting Review* 86 (3): forthcoming.

———, R. Hoitash, and U. Hoitash. 2009. Evidence from the United States on the effect of auditor involvement in assessing internal control over financial reporting. *International Journal of Auditing* 13 (2): 105–125.

Berry, L. E., G. B. Harwood, and J. L. Katz. 1987. Performance of auditing procedures by governmental auditors: Some preliminary evidence. *The Accounting Review* 58 (1): 14–28.

Brown, P. R. 1983. Independent auditor judgment in the evaluation of internal audit functions. *Journal of Accounting Research* 21 (2): 444–455.

———, and V. Karan. 1986. One approach for assessing the operational nature of auditing standards: An analysis of SAS 9. *Auditing: A Journal of Practice & Theory* 6 (1): 134–147.

Brown, L., and A. S. Pinello. 2007. To what extent does the financial reporting process curb earnings surprise games? *Journal of Accounting Research* 45 (5): 947–981.

Carcello, J., D. R. Hermanson, and K. Raghunandan. 2005. Factors associated with U.S. public companies investment in internal auditing. *Accounting Horizons* 19 (2): 69–84.

Clark, M. E., T. E. Gibbs, and R. G. Schroeder. 1980. Evaluating internal audit departments under SAS No. 9: Criteria for judging competence, objectivity, and performance. *The Woman CPA*. 22 (July): 8–11.

Church, B. K., and A. Schneider. 1995. Internal auditors' memory for financial-statement errors. *Behavioral Research in Accounting* 7: 17–36.

Davidson, R., J. Goodwin-Stewart, and P. Kent. 2005. Internal governance structures and earnings management. *Accounting and Finance* 45 (2): 241–267.

Doyle, J., W. Ge, and S. McVay. 2007a. Determinants of weaknesses in internal control over financial reporting. *Journal of Accounting and Economics* 44 (1–2): 193–223.

———, ———, and ———. 2007b. Accruals quality and internal control over financial reporting. *The Accounting Review* 82 (5): 1141–1170.

Ettredge, M. L., C. Li, and L. Sun. 2006. The Impact of SOX Section 404 Internal Control Quality Assessment on Audit Delay in the SOX Era. *Auditing: A Journal of Practice & Theory* 25 (2): 1–23.

Frankel, R. M., M. F. Johnson, and K. K. Nelson. 2002. The relation between auditors' fees for nonaudit services and earnings management. *The Accounting Review* 77 (s-1): 71–105.

Ge, W., and S. McVay. 2005. The disclosure of material weaknesses in internal control after the Sarbanes-Oxley Act. *Accounting Horizons* 19: 137–158.

Gompers, P., J. Ishii, and A. Metrick. 2003. Corporate governance and equity prices. *The Quarterly Journal of Economics* 118 (1): 107–155.

Goodwin, J., and P. Kent. 2006. Factors affecting the voluntary use of internal audit. *Managerial Auditing Journal* 21 (1): 81–101.

Gramling, A. A., M. J. Maletta, A. Schneider, and B. K. Church. 2004. The role of the internal audit function in corporate governance: A synthesis of the extant internal auditing literature and directions for future research. *Journal of Accounting Literature* 23: 194–244.

Harrell, A., M. Taylor, and E. Chewning. 1989. An examination of management's ability to bias the professional objectivity of internal auditors. *Accounting, Organizations and Society* 14 (3): 259–269.

Hogan, C. E., and M. S. Wilkins. 2008. Evidence on the audit risk model: Do auditors increase audit fees in the presence of internal control deficiencies?. *Contemporary Accounting Research* 25 (1): 219–242.

Hoitash, U., R. Hoitash, and J. C. Bedard. 2009. Corporate governance and internal control over financial reporting: A comparison of regulatory regimes. *The Accounting Review* 84 (3): 839–867.

Institute of Internal Auditors (IIA). 2002. *Practice Advisory 1110-2: Chief Audit Executive (CAE) Reporting Lines*. Available at: http://som.csudh.edu/depts/adjunct/mabohebeish/ACC%20336/PA%20ADVISORY/Practice_Advisory_1110-2.pdf.

———. 2004. *Internal Auditing's Role in Sections 302 and 404 of the Sarbanes-Oxley Act*. Available at: http://www.theiia.org/iia/download.cfm?file=1655.

———. 2008. *International Standards for the Professional Practice of Internal Auditing*. Available at: http://www.theiia.org/guidance/standards-and-guidance/ippf/standards/.

———. 2009a. *Formulating and Expressing Internal Audit Opinions*. Available at: http://www.theiia.org/guidance/standards-and-guidance/ippf/practice-guides/.

———. 2009b. *Practice Advisory 1110-1: Organizational Independence*. Available at: http://som.csudh.edu/depts/adjunct/mabohebeish/ACC%20336/PA%20ADVISORY/Practice_Advisory_1110-1.pdf.

Krishnan, J. 2005. Audit committee quality and internal control: An empirical analysis. *The Accounting Review* 80 (2): 649–675.

Krishnan, G. V., and G. Vishwanathan. 2007. Reporting internal control deficiencies in the post-Sarbanes-Oxley Era: The role of auditors and corporate governance. *International Journal of Auditing* 11: 73–90.

Kinney, W. 2000. Research opportunities in internal control quality and quality assurance. *Auditing: A Journal of Practice & Theory* 19 (Winter): 83–90.

Maletta, M. J. 1993. An examination of auditors' decisions to use internal auditors as assistants: The effect of inherent risk. *Contemporary Accounting Research* 9 (1): 508–525.

McDaniel, L. S. 1990. The effects of time pressure and audit program structure on audit performance. *Journal of Accounting Research* 28 (2): 267–285.

Messier, W. F., Jr., and A. Schneider. 1988. A hierarchical approach to the external auditors' evaluation of the internal auditing function. *Contemporary Accounting Research* 4 (2): 337–353.

New York Stock Exchange (NYSE). 2009. *Listed Company Manual*. Available at: http://nysemanual.nyse.com/lcm/.

PricewaterhouseCoopers (PwC). 2006. *2006 State of the Auditing Profession Study: Continuous auditing gains momentum*. Available at: http://www.pwc.com/us/en/internal-audit/publications/2009-study-internal-audit-profession.jhtml.

Public Company Accounting Oversight Board (PCAOB). 2004. *An Audit of Internal Control Over Financial*

Reporting Performed in Conjunction With an Audit of Financial Statements. Auditing Standard No. 2. Washington, D.C. Available at: http://www.pcaobus.org/Rules/Rules_of_the_Board/Auditing_Standard_2.pdf.

———. 2005. *Report on the Initial Implementation of Auditing Standard No. 2, An Audit of Internal Control Over Financial Reporting Performed in Conjunction with an Audit of Financial Statements*. Release No. 2005-023. November 30. Washington, D.C. Available at: http://pcaobus.org/Inspections/Pages/PublicReports.aspx.

———. 2007a. *An Audit of Internal Control Over Financial Reporting That Is Integrated with An Audit of Financial Statements*. Auditing Standard No. 5. Washington, D.C. Available at: http://www.pcaobus.org/Standards/Auditing/Pages/Auditing_Standard_5.aspx.

———. 2007b. *Report on the Second-Year Implementation of Auditing Standard No. 2, An Audit of Internal Control Over Financial Reporting Performed in Conjunction with an Audit of Financial Statements*. Release No. 2007-004. April 18. Washington, D.C. Available at: http://pcaobus.org/Inspections/Pages/PublicReports.aspx.

Prawitt, D. F., J. L. Smith, and D. A. Wood. 2009. Internal audit function quality and earnings management. *The Accounting Review* 84 (4): 1255–1280.

Rogers, J. L., and P. C. Stocken. 2005. Credibility of management forecasts. *The Accounting Review* 80 (4): 1233–1260.

Schneider, A. 1984. Modeling external auditors' evaluations of internal auditing. *Journal of Accounting Research* 22 (2): 657–678.

———. 1985. The reliance of external auditors on the internal audit function. *Journal of Accounting Research* 23 (2): 911–919.

———, and N. Wilner. 1990. A test of audit deterrent to financial reporting irregularities using the randomized response technique. *The Accounting Review* 65 (3): 668–681.

Schwartz, S. T., and R. A. Young. 2002. A laboratory investigation of verification and reputation formation in a repeated joint investment setting. *Contemporary Accounting Research* 19 (2): 311–342.

Securities and Exchange Commission (SEC). 2007. *Commission Guidance Regarding Management's Report on Internal Control over Financial Reporting Under Section 13(a) or 15(d) of the Securities Exchange Act of 1934*. Release Nos. 33-8810, 34-55929. Washington, D.C.: SEC.

Shumway, T. 2001. Forecasting bankruptcy more accurately: A simple hazard model. *Journal of Business* 74 (1): 101–125.

Stephens, N. M. 2009. Corporate governance quality and internal control reporting under SOX Section 302. Working paper, Utah State University.

U.S. Congress. 2002. The Sarbanes-Oxley Act of 2002. Public Law 107–2004. H.R. 3763. Washington, D.C.: Government Printing Office.

Wallace, W., and R. Kreutzfeldt. 1991. Distinctive characteristics of entities with an internal audit department and the association of the quality of such departments with errors. *Contemporary Accounting Research* 7 (2): 485–512.

Wright, A., and R. H. Ashton. 1989. Identifying audit adjustments with attention-directing procedures. *The Accounting Review* 64 (4): 710–728.

THE ACCOUNTING REVIEW
Vol. 86, No. 1
2011
pp. 325–357

American Accounting Association
DOI: 10.2308/accr.00000012

The Causes and Consequences of Internal Control Problems in Nonprofit Organizations

Christine Petrovits
The George Washington University

Catherine Shakespeare
University of Michigan

Aimee Shih
New York University

ABSTRACT: This study examines the causes and consequences of internal control deficiencies in the nonprofit sector using a sample of 27,495 public charities from 1999 to 2007. We first document that the likelihood of reporting an internal control problem increases for nonprofit organizations that are in poor financial health, growing, more complex, and/or smaller. We then present evidence that the disclosure of weak internal controls over financial reporting is negatively associated with subsequent donor support received after controlling for the current level of donor support and other factors influencing donations. We likewise report a negative association between internal control problems and subsequent government grants. Our results suggest that donors and government agencies, important sources of capital for nonprofit organizations, react either directly or indirectly to internal control information.

Keywords: *internal control; nonprofit organizations; donors; government grants; Sarbanes-Oxley.*

Data Availability: *Data are available from sources cited in the text.*

I. INTRODUCTION

The nonprofit sector represents a sizable slice of the United States economy. Nonprofit organizations had over $3.4 trillion in assets under their control and charitable giving to these organizations reached an estimated $295 billion, or 2.2 percent of gross domestic product, in 2006 (Wing et al. 2008). Several recent financial scandals have highlighted the sig-

We thank Phomika Palmer and Tom Pollak at the National Center for Charitable Statistics for help with data and Guidestar for providing access to historical Form 990s. We also thank Bill Baber, Mary Billings, Teresa Gordon, Angela Gore, Erica Harris, Steve Kachelmeier (editor), Irene Kim, Sarah McVay, Dan Tinkelman, Thomas Vermeer, Michelle Yetman, workshop participants at the 2007 ARNOVA Conference, the 2008 European Accounting Association Annual Congress, the 2008 American Accounting Association Annual Meeting, the 2009 FEA Conference, University of California, Irvine, Johns Hopkins University, Georgetown University, The George Washington University, and two anonymous reviewers for helpful comments.

Editor's note: Accepted by Steven Kachelmeier

Submitted: April 2009
Accepted: July 2010
Published Online: January 2011

nificant fiduciary responsibilities of nonprofit managers as well as the relatively weak regulatory oversight of the nonprofit sector.[1] As a result, lawmakers have increased calls for nonprofit organizations to adopt more rigorous corporate governance practices, including improved internal control practices.

Internal control audits are not new to the nonprofit sector. Nonprofit organizations that receive federal funding have been subject to reviews of internal control since 1990. We make use of this unique setting to investigate the causes of internal control deficiencies and perhaps, more interestingly, the consequences of internal control reporting for these organizations. Specifically, we examine the characteristics of public charities that report internal control problems and the effect of such problems on subsequent contributions and government grants received.

Internal control is broadly defined as the process put in place by management to provide reasonable assurance regarding the achievement of effective and efficient operations, reliable financial reporting, and compliance with laws and regulations. Thus, results of internal control audits provide information on the level of risk that a nonprofit organization is not effectively carrying out its mission-related activities and fiduciary responsibilities. For this study, we define an internal control problem as the existence of a reportable condition over financial reporting or over compliance with federal program requirements.

We first model the probability of disclosing an internal control problem as a function of salient characteristics of nonprofit organizations using a sample of 27,495 public charities from 1999 to 2007. Our results generally suggest that nonprofit organizations that are more complex, financially distressed, smaller, and/or growing rapidly are more likely to disclose an internal control problem, consistent with prior research (Ge and McVay 2005; Keating et al. 2005; Doyle et al. 2007a; Ashbaugh-Skaife et al. 2007).

Next, we consider the consequences of disclosure of an internal control problem for nonprofit organizations. Previous research into the consequences of an internal control deficiency focuses predominately on for-profit firms' cost of equity capital, either directly or indirectly through the market's response to the announcement of an internal control problem. The results from this research are mixed. Ashbaugh-Skaife et al. (2009) report that the disclosure of an internal control problem is associated with a higher cost of capital. However, using a different specification, Ogneva et al. (2007) find no relation between internal control deficiencies and the cost of equity capital. Furthermore, there is mixed evidence on the market response to internal control problems, with Section 302 disclosures generating negative abnormal returns but Section 404 disclosures having no effect (Beneish et al. 2008; Hammersley et al. 2008).

Nonprofit organizations do not issue shares and their missions are not to maximize profit. While nonprofit managers are not accountable to shareholders, they are accountable to donors and grantors who provide an important source of capital. These donors and grantors do not have limitless resources and, therefore, nonprofit organizations must compete for funding. If an organization reports an internal control problem, donors could choose to support another organization where, presumably, the capital is used more efficiently. Therefore, disclosure of an internal control deficiency could result in lower subsequent contributions. Alternatively, unlike shareholders, donors do not ultimately benefit from a nonprofit organization's activities and, thus, are less likely to monitor the organization (Fama and Jensen 1983). Some donors may be unaware of the problem or may not care about the problem and, therefore, the disclosure of an internal control deficiency may be unrelated to subsequent contributions.

---

[1] These scandals include the conviction of the CEO of the United Way of America for fraud; the Ponzi scheme perpetuated by the Baptist Foundation of Arizona, which an audit by Arthur Anderson failed to detect, that resulted in the largest nonprofit bankruptcy ever; the embezzlement of funds from ACORN by the founder's brother; and the lavish spending of university money by the president of Oral Roberts University, to name a few.

We examine whether the disclosure of an internal control problem is associated with lower contributions received subsequently from donors using the Weisbrod and Dominguez (1986) model, which captures the responsiveness of donations to various economic factors. We use a two-stage estimation procedure to control for endogeneity between internal control problems and contributions received. Our results indicate that reportable conditions over financial reporting are negatively associated with future public support, even after controlling for the current level of public support and other drivers of contributions. Organizations that disclose internal control problems over financial reporting receive fewer contributions from individuals, corporations, and foundations in the subsequent year.

Next, we investigate the effect of disclosing internal controls deficiencies on subsequent contributions received from local, state, and federal government agencies. Because audits are mandated by the federal government for recipients of federal funding and the results of these audits are filed with the Federal Audit Clearinghouse, governmental agencies likely use the information contained in the audit reports as one factor in funding decisions. Our results are consistent with expectations. We report negative associations between reportable conditions over both financial reporting and federal program compliance and subsequent government contributions, after controlling for prior-year government contributions and political and economic determinants of governmental funding allocations.

This study informs the debate over whether public charities should adopt more rigorous corporate governance practices, particularly in relation to internal control. Recently, policymakers have focused attention on the perceived lack of accountability and transparency by charitable organizations. This increased scrutiny is not necessarily unwarranted due to the recent financial scandals and the size of the nonprofit sector. Recognizing that they must maintain the public's trust, charities are working together to convince policymakers that they can address their shortcomings without onerous regulations.[2] However, the nonprofit sector has not focused much attention on the particular issue of internal control. Opponents of increased regulation argue that most donors do not use detailed financial information to make giving decisions and that nonprofits do not have the funds to comply with burdensome rules (e.g., Irvin 2005; Mulligan 2007). Our evidence suggests that the internal control information currently produced by a subset of organizations in the nonprofit sector does affect, either directly or indirectly, both donors' and government agencies' funding decisions.

The results of this study should also interest nonprofit managers who make decisions about how to allocate scarce resources. During difficult economic times, when the demand for services is skyrocketing, it is essential that nonprofit organizations continue to attract donors and grantors. These organizations face tremendous pressure to focus resources on mission-related activities. However, Hager et al. (2004) argue that pressure from donors and watchdog groups to maximize mission-related spending and limit overhead costs to artificially low levels is detrimental in the long-run. Our results are consistent with the Hager et al. (2004) argument that underinvestment in administrative expenses (e.g., internal controls) can ultimately have negative consequences on mission-related activities. In particular, our evidence suggests that improving internal controls not only reduces the risk of monetary loss resulting from fraud or accounting error, but may also increase the organization's ability to deliver services by attracting additional funding.

Furthermore, this study contributes to the literature on the consequences of internal control reporting as it provides a more direct measure of the response to internal control problems. Prior

---

[2] The most prominent example of self-regulation is the National Panel on the Nonprofit Sector convened by the Independent Sector. This panel proposed extensive changes in nonprofit governance and oversight in a June 2005 report to Congress, "Strengthening Transparency, Governance, and Accountability of Charitable Organizations."

research examines the impact on the cost of equity capital, which could be considered a less direct measure of stakeholder response than donor contributions and government grants, in part because it is inferred from market models under some potentially strong assumptions. In this study, we measure stakeholder response to internal control problems by investigating the change in donor and government support.

Finally, understanding the effects of disclosure of internal control problems is important because auditors of nonprofit organizations adopted SAS No. 112, *Communicating Internal Control Matters Identified in an Audit*, in 2007 and its successor SAS No. 115 of the same title in 2009. These standards define the types of internal control deficiencies, provide detailed guidance on evaluating the severity of internal control deficiencies, and require auditors to communicate in writing to management and those charged with governance any deficiencies noted in an audit (*Professional Standards*, AICPA 2010, vol. 1, AU §352.01). As a result, these standards may influence public perception of nonprofit organizations. For example, PricewaterhouseCoopers (2006, 2) notes, "If an auditor identifies an internal control issue, it must be reported to trustees, granting agencies, and other regulators under new definitions and in a more public manner than before and, as a result, control deficiencies could be exposed to greater scrutiny by stakeholders." Thus, consequences from reporting an internal control deficiency during the sample period are likely to be amplified under today's standards.

The next section outlines current nonprofit regulatory oversight, with an emphasis on internal control reporting. Section III presents our hypotheses and related empirical models. Section IV describes the sample selection procedures and data. Section V reports our results. Section VI concludes.

II. BACKGROUND

The nonprofit sector is growing rapidly in size and complexity. Approximately 1.4 million nonprofit organizations operate in the United States today (Wing et al. 2008). These organizations vary significantly in terms of mission, size, and primary revenue source. The Internal Revenue Code defines over 25 categories of nonprofits, such as human service organizations, schools, health care providers, cultural institutions, community development corporations, affordable housing, and research laboratories. Nonprofits exist to provide a public benefit, and, therefore, receive preferential tax treatment and other regulatory privileges. Most nonprofit organizations are either public charities or private charitable foundations organized under Section 501(c)(3) of the Internal Revenue Code.[3] Brown (2007) reports that, since 1997, the IRS has added to its master-file on average 39,465 exempt organizations per year, or 108 exempt organizations per day.

Regulatory oversight has not kept pace with the growth in the number of nonprofit organizations. Currently there are two regulatory mechanisms by which most nonprofit organizations are monitored: (1) the IRS via the organization's tax return (Form 990), which is required for all organizations receiving at least $25,000 of public support, and (2) the nonprofit laws in the state of incorporation, which vary widely from state to state. These mechanisms have been criticized as insufficient to ensure that nonprofits meet their fiduciary obligations (Hansmann 1981; Atkinson 1998; Fishman 2003; Reiser 2005). In fact, the IRS acknowledges a lack of enforcement presence (Brown 2007). Other monitoring mechanisms do exist but vary by the type of organization and

---

[3] Private foundations generally receive funding from a single source (i.e., a family or corporation), earn significant investment income, and make grants to other organizations. Public charities, as defined in Section 509(a), receive substantial support from the general public or government and actively conduct charitable operations. Private foundations are subject to various excise taxes and restrictions in order to ensure that they are using their resources for charitable purposes. Congress did not impose the same excise taxes and restrictions on public charities, presumably because donors hold public charities accountable.

type of funding sources (e.g., watchdog groups like the Better Business Bureau, program service contracts like Medicare, periodic program evaluations required by foundation and corporate donors, and state telemarketing reporting).

Congress passed the Sarbanes-Oxley Act of 2002 in an attempt to improve the accountability and oversight of public companies. Most of the provisions of the Sarbanes-Oxley Act do not apply to nonprofit organizations and no federal equivalent of the Act currently exists for nonprofits.[4] Nevertheless, Sarbanes-Oxley influences attitudes about corporate governance in the nonprofit community (Ostrower 2007; Iyer and Watkins 2008). Policymakers at both the state and federal levels are considering various proposals aimed at enhancing nonprofit accountability (Fremont-Smith 2007). For example, Senator Charles Grassley (2006, 26), then Chairman of the Senate Finance committee, said:

> Just as Congress has acted in the public interest to protect shareholders and workers from corporate mismanagement, so too must Congress demand transparency, accountability, and good governance from the nonprofit sector… Tightening rules and regulations governing the nonprofit sector will help repair the breach of trust that threatens to tarnish even the most reputable charities in America.

One of the main elements of Sarbanes-Oxley is management's responsibility for internal controls. Section 302 of the Act requires that chief executive and chief financial officers evaluate the design and effectiveness of internal controls on a quarterly basis and report an overall conclusion about the effectiveness of internal controls. Section 404 of the Act requires an annual audit of management's evaluation of internal controls and of the effectiveness of internal controls. Even though public charities are not subject to either Section 302 or Section 404 of Sarbanes-Oxley, similar requirements have been considered for the nonprofit sector. For example, the attorneys general in the states of New York and Massachusetts have proposed bills with provisions similar to the requirements in Section 302.

Some charities already are required to undergo internal control evaluations annually because they receive federal funding. Specifically, all organizations with federal expenditures greater than $500,000 ($300,000 for fiscal years ending before January 1, 2004) must have an audit conducted in accordance with Office of Management and Budget (OMB) Circular A-133 "Audits of States, Local Governments and Non-Profit Organizations."[5] The results of these audits (Form SF-SAC) must be filed within nine months of the end of the fiscal year with the Federal Audit Clearinghouse and are publicly available.

The objective of an A-133 audit, also called a single audit, is to provide assurance that an organization receiving grants from the federal government is using the funds appropriately and is complying with all federal regulations (AICPA 2009). As part of an A-133 audit, auditors issue opinions on both the financial statements and on compliance with the provisions of the federal contracts or grants. In addition, auditors report on internal control over both financial reporting and federal program compliance.[6] The reports on internal control identify whether there are any reportable conditions and, if so, whether any reportable conditions are material weaknesses.[7]

---

[4] The two provisions of SOX that do explicitly apply to nonprofit organizations are whistle blower protection and document destruction policies.

[5] OMB Circular A-133 was issued in 1990 under the name "Audits of Institutions of Higher Education and Other Non-Profit Institutions" and revised in 1996, 2003, and 2007. Our sample includes observations before and after the 2003 revision. The 2003 revision includes raising the audit threshold to $500,000 and technical changes related to the determination of cognizant agency (Federal Register 68 (June 27, 2003): 38401-38402). We include year controls in our empirical tests to control for any differences across time.

[6] Technically, auditors must test compliance requirements for "major programs." The determination of major programs takes into account risk, size, and oversight by a federal agency. See AICPA (2003) for more details.

[7] Reportable conditions involve deficiencies in the design or operation of internal controls that could adversely affect the organization's financial reporting or its ability to administer its federal programs. Material weaknesses are reportable

Circular A-133 establishes certain conditions for determining whether a nonprofit organization qualifies as a low-risk auditee. This determination is based on numerous factors including prior-year audit results, third-party references, the level of oversight of the granting federal agency, and the inherent risk of the federal programs involved. To be considered low-risk, an organization must have been audited annually for the past two years and these prior audits must have resulted in clean opinions, no internal control deficiencies, and no audit findings. The risk determination affects the amount of auditing that is required to be performed under OMB A-133. For organizations not deemed low-risk, auditors are required to perform more testing and, thus, are more likely to uncover internal control problems.

The assessment of internal controls required by Circular A-133 is not identical to the assessment of internal controls required by Sarbanes-Oxley. The A-133 audit is not overseen by the PCAOB, but rather is performed in accordance with *Government Auditing Standards* issued by the GAO. As noted above, the scope of an A-133 audit goes beyond financial statements to include the requirements of federal grants or contracts. Although the scope is wider, A-133 audits are generally less stringent and less costly. Despite the differences from public company audits, A-133 audits do provide information on the level of risk that a nonprofit organization is not effectively carrying out its mission-related activities and fiduciary responsibilities.

III. HYPOTHESIS DEVELOPMENT AND EMPIRICAL MODELS

Determinants of Internal Control Deficiencies

We first examine the determinants of internal control deficiencies. A significant body of research examines the characteristics of publicly traded companies disclosing internal control problems. Ge and McVay (2005) find that firms disclosing material weaknesses are more complex, smaller, and less profitable than firms not disclosing material weaknesses. Doyle et al. (2007a) add that firms disclosing material weaknesses are younger, growing rapidly, or undergoing restructuring. Likewise, Ashbaugh-Skaife et al. (2007) find that firms reporting internal control deficiencies have more complex operations, greater exposure to accounting risk, fewer resources to invest in internal control, and a higher likelihood of using a dominant auditor.

Despite the extensive academic literature on internal controls in publicly traded companies, there is little research on internal control in the nonprofit sector. Keating et al. (2005) examine A-133 audit results from 1997 to 1999 using univariate tests. They find that smaller organizations and organizations not classified as low-risk (i.e., new grantees, organizations with prior problems) disclose more internal control problems. They also report that organizations with audits performed by national, large regional, and specialist firms report fewer internal control problems, which differs from the Ashbaugh-Skaife et al. (2007) auditor quality results for public companies. Keating et al. (2005) suggest that small nonprofit organizations, which are more likely to have internal control problems, select small audit firms.

We extend Keating et al. (2005) by examining a more comprehensive set of factors that may be associated with reporting internal control deficiencies in nonprofit organizations. Specifically, we model the likelihood of reporting internal control problems as a function of several internal control risk factors and audit detection variables. As discussed below, we expect organizations that are more complex, in poor financial health, smaller, new to federal funding, and/or growing rapidly to disclose more internal control deficiencies.

---

conditions in which the design or operation of internal controls does not reduce to a relatively low level the risk of material noncompliance with applicable grant requirements or with GAAP caused by error or fraud that may occur and not be detected in a timely manner (AICPA 2003, 104). OMB revised Circular A-133 in June 2007 to be consistent with SAS No. 112, which replaced the "reportable condition" concept with "significant deficiency." Because our sample period pre-dates SAS No. 112, we use the term reportable condition.

Public charities with diverse operations face challenges instituting internal controls across their various initiatives and divisions. We measure organizational complexity using the number of revenue sources (public support, government support, and/or program service revenue.) Organizations that receive funding from only one source generally engage in fewer types of charitable programs than organizations that receive funding from several sources. We predict that organizations with more sources of funding (*COMPLEXITY*) are more likely to have a variety of operations and, therefore, more likely to report an internal control problem.

Nonprofit organizations in poor financial health are less likely to have resources to invest in establishing strong internal controls. We use the existence of a going-concern paragraph in the opinion on the financial statements (*GOINGCONCERNRISK*) as a proxy for poor financial health. A going-concern paragraph indicates that the auditor has substantial doubt whether the organization can meet its obligations as they become due. We expect that organizations with a going-concern paragraph report more internal control deficiencies. Consistent with studies of public companies that use the existence of losses to measure financial health, we also include an indicator of whether the organization's revenues exceed its expenses (*SURPLUS*). We expect that charities with a surplus have fewer internal control problems.

Larger organizations (*SIZE*), as measured by total assets, have more resources and experience to draw on when implementing internal controls. For example, Greenlee et al. (2007) report that older and larger nonprofit organizations are more likely to have an internal audit function in place. Thus, we expect that smaller organizations disclose more internal control problems. Internal controls should change in response to organizational change as existing controls may be irrelevant or inefficient and new controls may be required. Rapidly growing organizations are often unable to adequately assess and update internal controls at the same pace at which organizational expansion occurs. We predict that change in size (*GROWTH*) is positively associated with the existence of internal control deficiencies. Similarly, organizations that receive federal funding for the first time (*NEWGRANTEE*) are less likely to have all of the internal controls systems in place to meet federal requirements.

We also investigate the effect of auditor type on the probability of reporting an internal control problem but do not make a prediction. On one hand, dominant audit firms (*BIG4*, *REGIONAL*, and *SPECIALIST*) have more training, experience, and exposure to litigation risk, all of which imply that these audit firms are more likely to discover internal control deficiencies. On the other hand, dominant audit firms may only contract with prestigious nonprofit organizations that are inherently less risky. This self-selection suggests that dominant audit firms are less likely to discover internal control problems at their nonprofit clients. In fact, Kitching (2009) finds evidence that donors behave as if dominant auditors are a signal of credibility.

As noted in Section II, auditors are required to determine whether a nonprofit organization qualifies as a low-risk auditee under OMB A-133. Because they are inherently less risky and because there is less testing involved, the likelihood of detecting an internal control problem is lower for low-risk auditees than it is for other auditees. Thus, we include an indicator variable if the organization is not deemed low-risk (*RISK*) as an important control in our model.

Based on the above discussion, we estimate the probability of disclosing an internal control deficiency as a function of organizational characteristics and audit detection variables as follows:

$$\begin{aligned} Prob(ICD) = \beta_{0i} + \beta_1 COMPLEXITY + \beta_2 GOINGCONCERNRISK + \beta_3 SURPLUS + \beta_4 lnSIZE \\ + \beta_5 GROWTH + \beta_6 RISK + \beta_7 NEWGRANTEE + \beta_8 BIG4 + \beta_9 REGIONAL \\ + \beta_{10} SPECIALIST + \Sigma \gamma_i INDUSTRY + \Sigma \delta_i YEAR. \end{aligned} \tag{1}$$

Overall, we expect that the likelihood of reporting an internal control problem increases as a function of *COMPLEXITY*, *GOINGCONCERNRISK*, *GROWTH*, *RISK*, and *NEWGRANTEE* and

decreases as a function of *SURPLUS* and *SIZE*. The empirical specification also includes controls for industry and year.

Effect of Internal Control Deficiencies

We next examine the consequences of internal control deficiencies. Prior studies of public companies document that internal control problems are associated with equity market concerns. Specifically, firms reporting an internal control deficiency under Section 302 experience stock price declines, with the most negative returns for material weakness disclosures (Hammersley et al. 2008; Beneish et al. 2008). Evidence on the impact of internal control problems on the cost of capital for public companies is mixed. Ashbaugh-Skaife et al. (2009) find that internal control problems are associated with a higher cost of equity, while Ogneva et al. (2007), using a different specification, do not find an association. A limitation of these studies of public companies is that the cost of equity is a less direct measure of stakeholders' reactions, as it is inferred from a market model under certain strong assumptions.

There has been little consideration given to understanding the effect of internal control deficiencies in the nonprofit sector. Internal controls are established to provide assurance that operations are running efficiently and that financial reporting is reliable. We expect that nonprofit organizations with internal control problems have lower operating efficiency and produce lower quality financial reports, on average. Thus, internal control problems can influence directly or indirectly the amount of funds available to achieve the organization's mission. The source of funding takes many forms depending on the type of organization, including donor contributions, government grants, program service revenue, and/or debt financing. As discussed below, we examine the impact of internal control problems on donations (*PUBLIC SUPPORT*) and government grants (*GOV CONTRIBUTIONS*). Unlike the indirect cost of capital measure for public companies noted above, contributions by donors and government agencies to nonprofit organizations provide direct evidence of stakeholder reactions to internal control problems.

Public Support

Public support includes gifts received from individuals, trusts and estates, corporations, and foundations (*DIRECT SUPPORT*), as well as gifts received from federated fundraising agencies (*INDIRECT SUPPORT*), such as the United Way and the Combined Federal Campaign. Donors generally have less information about the quality of the nonprofit organization's output relative to government grantors, customers (who provide program service revenue), and creditors. Nevertheless, donors provide a substantial amount of support to the nonprofit sector. In the face of this information asymmetry, it is important to understand all factors, including the quality of internal control, that influence a donor's charitable giving decision in a competitive market for donations.

Several prior studies offer evidence that a public charity's operating efficiency is positively associated with the amount of donor support received (e.g., Weisbrod and Dominguez 1986; Posnett and Sandler 1989; Greenlee and Brown 1999; Tinkelman 2004; Tinkelman and Mankaney 2007).[8] Further, Yetman (2008) reports that donors give less to organizations that overstate mission-related expenses and understate fundraising expenses, providing some support for the idea that donors can unravel low-quality financial statements. However, Tinkelman (1998) and Khumawala et al. (2005) both provide evidence that most donors do not unravel joint cost allocations made to strategically overstate mission-related activities. Overall, prior research suggests that, in

[8] There are many slight variations in the definition of operating efficiency. In general, these operating efficiency measures attempt to capture how much the nonprofit organization spends on program-related activities (i.e., fulfilling its mission) relative to how much it spends on administrative and fundraising costs.

many but not all cases, donors use available information from the organization's Form 990 to distinguish higher quality nonprofit organizations from lower quality nonprofit organizations and make their giving decisions accordingly.

Because internal control deficiencies can signal a lack of effectiveness in providing charitable services and a higher probability of undetected misconduct, all else equal, we expect that nonprofit organizations with internal control deficiencies receive fewer subsequent contributions from the public than organizations with no internal control deficiencies. This hypothesis is based on the assumption that donors make giving decisions in order to assist in the provision of public goods and, thus, opt to give to organizations that can provide the public goods with minimum risks. While it is likely that some financially sophisticated donors (e.g., private foundations) actually obtain the publicly available A-133 audit report as part of the giving decision process, it is highly unlikely that all donors do. Even if donors do not directly learn that an organization has reported an internal control problem from the A-133 report, they may still indirectly receive information about that problem. For example, internal control problems can be associated with lower operating efficiency, which is observable on the more widely distributed Form 990. Alternatively, a donor may have lower quality interactions with an organization that has internal control problems (e.g., an internal control weakness causes donor acknowledgments not to be sent as required by the IRS).

There are reasons why the quality of an organization's internal controls may not affect public support. In particular, not all donors give in order to provide a public good. Some donors simply seek a warm glow (Andreoni 1990; Ribar and Wilhelm 2002) and, thus, internal control information, or any financial information for that matter, is irrelevant. Also, it may be too costly for donors to obtain and evaluate A-133 audit information.[9] Finally, if the internal control audit results are not filed until nine months after year-end, the information may be stale. Therefore, it is an open empirical question as to whether internal control problems affect subsequent contributions.

We adapt the widely used Weisbrod and Dominguez (1986) approach to capture the responsiveness of donations to various economic factors. Weisbrod and Dominguez (1986) model public support as a function of conventional market variables, including price, fundraising expenses, and age:

$$lnPUBLIC\ SUPPORT_t = \beta_0 + \beta_1 lnFUNDRAISING\ EXP_{t-1} + \beta_2 lnPRICE_{t-1} + \beta_3 AGE_t.$$

PRICE measures the cost to the donor of "purchasing" (i.e., contributing) one more dollar of the organization's charitable output. *PRICE* depends on the after-tax cost of giving, as well as the efficiency by which the organization generates output. Specifically, *PRICE* is defined as:

$$PRICE = \frac{(1-T)}{1-\left(\dfrac{FUNDRAISING\ EXP + ADMINISTRATIVE\ EXP}{TOTAL\ EXPENSE}\right)}.$$

Donors face the same marginal tax rate with respect to donations for all charitable organizations and, thus, we assume $T = 0$. Note that the denominator is equivalent to the program expense ratio (Program Expenses/Total Expenses) so when $T = 0$, *PRICE* equals the inverse of the program expense ratio (Total Expenses/Program Expenses).[10] Theoretically, price should have a negative

---

[9] This discussion suggests that any influence that internal control does have on public support is moderated by the level of sophistication of the organization's donor clientele. Unfortunately, it is impossible to test this supposition using archival data because nonprofit organizations do not disclose the identities of their donors in a consistent, systematic manner. See Baber et al. (2001) for a discussion of donor clienteles.

[10] When defining *PRICE*, some studies scale fundraising and administrative expenses by public support (e.g., Weisbrod

influence on the level of giving. However, Bowman (2006) notes that, in prior empirical studies, results of tests examining the effect of price on public support are sensitive to model specification. The Weisbrod and Dominguez (1986) model also includes *FUNDRAISING EXP*, which represents the organization's effort to reduce information asymmetry, and *AGE*, which represents the organization's stock of goodwill. Both are expected to positively affect public support.

In order to test our expectation about the effect of internal control problems on donor contributions, we estimate the following equation:

$$\begin{aligned} lnPUBLIC\ SUPPORT_t = {} & \beta_0 + \beta_1 INTERNAL\ CONTROL\ DEFICIENCY_{t-1} \\ & + \beta_2 lnFUNDRAISING\ EXP_{t-1} + \beta_3 lnPRICE_{t-1} + \beta_4 AGE_t \\ & + \beta_5 lnGOV\ CONTRIBUTIONS_{t-1} + \beta_6 lnPROGRAM\ REVENUE_{t-1} \\ & + \beta_7 lnPUBLIC\ SUPPORT_{t-1} + \Sigma\gamma_i INDUSTRY + \Sigma\delta_i YEAR. \end{aligned} \tag{2}$$

In addition to the Weisbrod and Dominguez (1986) variables, we include government grants (*GOV CONTRIBUTIONS*) and program service revenue (*PROGRAM REVENUE*) in order to control for any crowding-out or crowding-in effects. Khanna and Sandler (2000), and Okten and Weisbrod (2000) provide evidence of a positive relation between public support and government grants and program service revenue, indicating a crowding-in effect.[11] Finally, we include prior-year public support to capture any other organization-specific factors, as well as industry and year controls. We are primarily interested in the coefficient on internal control deficiency, β_1, and predict that the existence of an internal control problem is negatively associated with subsequent public support.[12]

Government Contributions

Government contributions represent gifts and grants from federal, state, and local governments.[13] Similar to our expectation for public support, we hypothesize that nonprofit organizations with internal control deficiencies receive fewer government contributions than organizations with no internal control deficiencies. Given that the federal government mandates internal control reporting as part of the required A-133 audit, all else equal, the federal government should use this internal control information to make funding decisions.

Even though our prediction of a negative association between internal control problems and subsequent government contributions seems intuitive, there are questions about the actual deter-

---

and Dominguez 1986; Okten and Weisbrod 2000; Bowman 2006), while others scale by total expenses (e.g., Posnett and Sandler 1989; Tinkelman 1998; Khanna and Sandler 2000). That is, some studies measure the amount of charitable output relative to revenue received and some measure charitable output relative to total expenses. Our results are not affected by the choice of scale. We choose to present the results scaling by total expenses for a practical reason. In many instances, the sum of fundraising and administrative expenses exceeds public support. In these instances, log(*PRICE*) is undefined. Thus, scaling by public support limits the number of usable observations.

11 Crowding-out occurs when an increase in government support discourages donors from increasing their own contributions because need is already being met by government support. Crowding-in occurs when government support encourages donors to increase their own giving, often because government grants enhance the reputation of the nonprofit organization and because government grants are accompanied by more monitoring that provides additional assurance to donors that their funds are being used appropriately.

12 We examine the association between internal control problems and public support received in the following year. To the extent that information from internal control audits is not available until nine months after the fiscal year-end and donors directly use this information, our tests are biased against finding results. Our approach is consistent with prior studies that examine the influence of program-spending ratios on subsequent years' giving using 990 data, where 990s are generally filed from 5 to 11 months after year-end.

13 It is important to note that government contributions on IRS Form 990 are distinct from government contracts or payments for service, which are included in program service revenue (e.g., Medicare payments received by a hospital are not classified as government contributions). Thus, it is common for organizations to receive no government contributions but still qualify for the A-133 audit because they earn revenue from federal contracts.

minants of government funding in the nonprofit sector. It is possible that political factors, and not the quality of the nonprofit organization as signaled by the internal control audit results, drive government contribution decisions. For example, the New York Attorney General recently initiated a probe into "pay-to-play" campaign donations made by nonprofit organizations to politicians in order for the nonprofit organizations to receive government grants (Dicker and Goldenberg 2009). In addition, government contributions include funds received from state and local governments, which may be less likely to use the A-133 audit report than the federal government.[14] Finally, the federal government comprises a wide variety of federal agencies. There are 55 different federal agencies with oversight responsibilities in our sample, ranging from the CIA to the Peace Corps and from the National Science Foundation to the National Endowment for the Arts.[15] These agencies have different missions and likely use internal control information differently. For these reasons, we empirically examine the link between internal control weaknesses and subsequent government grants.

We estimate a model of government contributions as a function of internal control problems and political and socio-economic factors, as follows:

$$
\begin{aligned}
lnGOV\ CONTRIBUTIONS_t = \beta_0 &+ \beta_1 INTERNAL\ CONTROL\ DEFICIENCY_{t-1} \\
&+ \beta_2 lnFUNDRAISING\ EXP_{t-1} + \beta_3 LOBBYING_t + \beta_4 AGE_t \\
&+ \beta_5 lnPRICE_{t-1} + \beta_6 lnPUBLIC\ SUPPORT_{t-1} \\
&+ \beta_7 lnPROGRAM\ REVENUE_{t-1} \\
&+ \beta_8 lnGOV\ CONTRIBUTIONS_{t-1} + \beta_9 GDP_t + \Sigma\alpha_i STATE \\
&+ \Sigma\gamma_i INDUSTRY. \qquad (3)
\end{aligned}
$$

In order to address the political determinants of government funding, we include *LOBBYING*, which is an indicator variable that designates whether the organization incurred expenditures to influence legislation through communication with legislators or government officials.[16] While lobbying may not, *per se*, result in more government funding, engaging in lobbying activities signifies a politically savvy nonprofit organization. Thus, we expect a positive association between lobbying and government contributions.

We include annual gross domestic product (*GDP*) to control for economic conditions.[17] Governments should have a greater supply of funds available for gifts and grants when *GDP* is higher. Alternatively, the demand for government grants is higher during periods of low *GDP*. Thus, we make no predictions about the coefficient on *GDP*.

We also use state indicator variables as proxies for demand for government funding. One objective of government contributions is to redistribute revenue to geographic areas with eco-

---

[14] Different states and municipalities have different levels of monitoring of their grant programs. It is not possible to identify the specific source(s) of government contributions from the Form 990. The largest portion of government contributions for most nonprofits comes from the federal government, although large, complex organizations receive contributions from numerous federal, state, and local agencies.

[15] Nonprofit organizations expending more than \$50 million (\$25 million for fiscal years ending before January 1, 2004) in federal awards are assigned a cognizant agency. All other nonprofits are assigned an oversight agency. Generally, the cognizant or oversight agency is the federal agency that provides the predominant amount of federal funding. The purpose of the cognizant or oversight agency is to provide technical audit advice and act as a liaison between the nonprofit organization and other federal agencies with respect to audit issues.

[16] Nonprofit organizations are not allowed to make political campaign donations but are allowed to engage in some lobbying activities, subject to certain limitations, without risking their tax-exempt status. See Treasury Regulations Section 56.4911 for more details.

[17] Because we measure GDP on an annual basis, we cannot concurrently include year controls in the model.

nomic need. In some cases, the level of funding is established mathematically by the population served by the nonprofit organization (e.g., need-based formula grants). While nonprofit organizations do not operate exclusively in one geographic region, the state in which these organizations are headquartered can generally reflect characteristics of the populations that they serve (i.e., level of poverty). State indicator variables may also control for political factors (e.g., state representation on Congressional appropriation committees).

Andreoni and Payne (2003) find that nonprofit organizations reduce fundraising efforts when they receive government grants. If a nonprofit organization determines that government support will be slashed in the next period, either because the organization has fallen out of political favor or because of macroeconomic constraints, the nonprofit organization may increase fundraising. Thus, we expect a negative coefficient on *FUNDRAISING EXP*.

Generally, individual donors use *PRICE* as a measure of operating efficiency, while governments have more direct methods of monitoring an organization's efficiency (Khanna and Sandler 2000). Nevertheless, some government agencies, particularly at the state and local level, could use basic operating efficiency ratios to make decisions and/or *PRICE* could serve as a proxy for the more complex efficiency measures actually used by government agencies. Thus, we include *PRICE*, as defined in the previous section, in Equation (3). We also include *AGE* as an indicator of reputation, *PUBLIC SUPPORT* and *PROGRAM REVENUE* as controls for any crowding-in or crowding-out effects, prior-year *GOV CONTRIBUTIONS* to capture any other organization-specific factors, and industry controls. As with Equation (2), we are primarily interested in the coefficient on internal control deficiency, β_1, and expect that the disclosure of an internal control problem is negatively associated with subsequent government contributions.

A final note—reporting a negative coefficient on β_1 in either Equation (2) or Equation (3) may indicate that disclosure of internal control deficiencies influences giving decisions. However, it is also possible that low levels of contributions result in inadequate resources necessary for a nonprofit organization to implement strong controls. To address endogeneity concerns, we implement the Heckman (1979) selection model. In the first stage, we use Equation (1) to estimate the likelihood of reporting an internal control deficiency and, using the parameters of this model, compute an inverse Mills ratio. In the second stage, we estimate Equation (2) and Equation (3) with the inverse Mills ratio as a control.

IV. SAMPLE SELECTION AND DATA

We obtain data on public charities from two sources: (1) the A-133 Single Audit database available from the Federal Audit Clearinghouse and (2) the IRS Form 990 databases available from the National Center for Charitable Statistics (NCCS). The A-133 data include general auditee information, the amount of federal awards expended, auditor name, type of audit performed, audit opinions, internal control information, and audit findings as reported on the Form SF-SAC.[18] The IRS data include revenues, expenses, and balance sheet data as reported on the Form 990.[19] All variables that we use from these databases are defined in Table 1.

---

[18] Information on internal controls over major programs is not available in the electronic Single Audit database until 2001 when the federal government changed the format of Form SF-SAC. Information on internal controls over financial reporting is available for all years.

[19] We use data from several different Form 990 databases. The Core Trend v2009a file provides organizational characteristics and basic financial statement data from 1998–2007. The DD Revenues and Expenses v2005 file provides a detailed breakdown of revenues and expense categories from 1998–2003. The SOI file provides a detailed breakdown of revenue and expense categories from 2004–2007 for a stratified random sample of firms selected by the IRS. The Core Supplemental v2009 file provides a detailed breakdown of revenues and expenses categories for 2004–2006 for organizations not covered in the SOI file. Note our sample does not include detailed data for the universe of organizations for 2007 but rather only for the sample in the SOI file. Our results are consistent when we exclude 2007 from our analysis.

TABLE 1

Variable Definitions

Variable	Definition
RC_FS	= An indicator variable that equals 1 if the A-133 audit noted reportable conditions in internal controls over financial reporting (SF-SAC Part II #3); otherwise 0.
RC_GOV	= An indicator variable that equals 1 if the A-133 audit noted reportable conditions in internal controls over major programs (SF-SAC Part III #5); otherwise 0.
RC_ANY	= An indicator variable that equals 1 if *RC_FS* equals 1 or *RC_GOV* equals 1; otherwise 0.
PUBLIC SUPPORT	= Total public support received for the fiscal year, defined as the sum of direct support and indirect support.
DIRECT SUPPORT	= Direct public support received for the fiscal year (Form 990 Line 1a).
INDIRECT SUPPORT	= Indirect public support received for the fiscal year (Form 990 Line 1b).
PROGRAM REVENUE	= Program service revenue, including government fees and contracts, received for the fiscal year (Form 990 Line 2).
GOV CONTRIBUTIONS	= Government contributions (grants) received for the fiscal year (Form 990 Line 1c).
FUNDRAISING EXP	= Fundraising expenses for the fiscal year (Form 990 Line 15).
COMPLEXITY	= Number of revenue sources included on Form 990 from 0–3 (*PUBLIC SUPPORT, GOV CONTRIBUTIONS*, and/or *PROGRAM REVENUE*)
GOINGCONCERNRISK	= An indicator variable that equals 1 if the A-133 audit includes a going-concern explanation (SF-SAC Part II #2); otherwise 0.
SURPLUS	= An indicator variable that equals 1 if total revenues (Form 990 Line 12) − total expenses (Form 990 Line 17) ≥0; otherwise 0.
NEWGRANTEE	= An indicator variable that equals 1 if the current year is the first year an organization expends federal contributions; otherwise 0.
AGE	= Number of years the organization has been tax-exempt (IRS RuleDate).
SIZE	= Beginning-of-year total assets (Form 990 Line 59a).
GROWTH	= The growth in assets, measured as the ratio of end-of-year total Assets (Form 990 Line 59b) to beginning-of-year total Assets (Form 990 Line 59a).
RISK	= An indicator variable that equals 1 if organization is classified as "not low risk" on the A-133 audit (SF-SAC Part III #3); otherwise 0.
BIG4	= An indicator variable that equals 1 if auditor (SF-SAC Part I #7) of the A-133 report is classified as one of the Big 6 auditors; otherwise 0.
REGIONAL	= An indicator variable that equals 1 if auditor (SF-SAC Part I #7) of A-133 report is classified as one of the Regional auditors; otherwise 0.

(continued on next page)

TABLE 1 (continued)

Variable	Definition
SPECIALIST	= An indicator variable that equals 1 if auditor (SF-SAC Part I #7) of the A-133 report is classified as one of the Specialist auditors; otherwise 0.
PRICE	= Total Expenses/Program Service Expense, where Total Expenses is the sum of fundraising expenses (Form 990 Line 15), management and general expenses (Form 990 Line 14), and program services expenses (Form 990 Line 13).
LOBBYING	= An indicator variable that equals 1 if organization reports lobbying expenditures to directly influence a legislative body (Form 990 Schedule A Line 37b); otherwise 0.

Source: IRS Form 990 from the National Center for Charitable Statistics and Form SF-SAC from the Federal Audit Clearinghouse.

Table 2, Panel A details the sample selection process. A merge of the A-133 data and the IRS Core Trend data on EIN and year results in a sample of 127,988 observations (27,495 public charities) from 1999 to 2007.[20] We use the Full Sample to shed light on the determinants of internal control weaknesses for a broad cross-section of organizations.

As discussed in the previous section, the second stage of our tests focuses on the influence of internal control problems on subsequent public support and government contributions. Some public charities receive only an immaterial amount of public support and/or government contributions (e.g., low-income housing projects). Thus, for the second stage, we further limit our sample to the subsets of organizations where public support or government contributions represent a nontrivial source of revenue. Specifically, following Tinkelman and Mankaney (2007), we eliminate observations with public support (government contributions) of less than $100,000. This process results in a Public Support Sample of 47,318 observations (12,342 public charities) and a Government Sample of 65,415 observations (16,369 public charities). These limited samples are still significantly larger than samples in previous studies of internal control in the for-profit literature. Lack of necessary data further reduces sample size for specific tests.

In Table 2, Panel B, we classify observations into five main industries based on the National Taxonomy of Exempt Entities (NTEE) developed by the IRS. The five industries, which are the same industries used in Keating et al. (2005), include: Arts, Education, Health, Human Services, and Public Benefit. The remaining NTEE categories (i.e., Religion, International, Environment, and Unknown) are classified as "Other." Human Services organizations (e.g., Red Cross chapters, YMCAs) comprise approximately half of the sample, while Arts and Cultural organizations comprise the smallest fraction of the sample. Untabulated results indicate that, although Human Services are the most common type of nonprofit organizations receiving federal awards, these organizations are also the smallest as measured by total assets. Educational institutions, which

[20] Because the NCCS data contain some data errors we conduct the error-checking procedures recommended by the NCCS. We noted some organizations that had identical information in consecutive years. We could not universally determine which year contained the correct information and, thus, deleted all related years. This resulted in a loss of 9,735 observations. Also, as suggested by the 2006 *Guide to Using NCCS Data*, any suspicious observations were compared to full text versions of the Form 990 available at Guidestar (http://www.guidestar.org). A small number of corrections were made, primarily related to the units reported (i.e., the file listed $5 instead of $5 million). To our knowledge, any remaining errors create noise but do not systematically bias our tests.

TABLE 2

Sample Description

Panel A: Sample Selection

	Full Sample	Public Support Sample	Government Sample
Public operating charities reporting to IRS	2,261,486		
Organizations receiving A-133 audit	337,353		
Merge IRS and A-133 data	129,356		
Less audit periods other than "Annual"	(1,368)		
Full Sample	127,988		
Less public (government) support$_{t-1}$ < \$100,000		(75,389)	(54,384)
Less no public (government) support$_t$ data		(5,281)	(8,189)
Total Observations	127,988	47,318	65,415
Unique Organizations	27,495	12,342	16,369

Panel B: Observations by NTEE Classification

	Full	Public Support	Government
Arts	1,912	1,136	1,268
Education	15,574	9,435	9,688
Health	26,397	10,397	14,935
Human Services	69,320	20,788	30,667
Public Benefit	10,923	3,745	6,761
Other	3,862	1,817	2,096
Total Observations	127,988	47,318	65,415

Panel C: Observations by Auditor Type

	Full	Public Support	Government
Big 4	11,254	7,370	6,690
Regional	11,467	5,225	6,051
Specialist	18,768	6,042	8,493
Other	86,499	28,681	44,181
Total Observations	127,988	47,318	65,415

NTEE classifications: Arts (Major Group A), Education (Major Group B), Health (Major Groups E, F, G, H), Human Services (Major Group I, J, K, L, M, N, O, P), and Public Benefit (Major Groups R, S, T, U, V, W).

comprise approximately 12 percent of the Full Sample and 20 percent of the Public Support Sample, are significantly larger than other types of nonprofit organizations. In Table 2, Panel C, we classify observations by the type of auditor, which is based on audit firm size and experience in conducting A-133 audits, similar to Keating et al. (2005). The Big 4 category includes the largest public accounting firms during the sample period.[21] The Regional category includes any of the next 25 largest accounting firms from *Accounting Today*'s 2004 Top 100 Firms list (ranked by

[21] This category includes any nonprofit organization audited by Deloitte & Touche, Ernst & Young, KPMG, PricewaterhouseCoopers (Coopers & Lybrand or Price Waterhouse), or Arthur Andersen.

revenues). The Specialist category includes accounting firms, not already classified as Big 4 or Regional, that conducted 100 or more A-133 audits during the sample period. The Other category includes all accounting firms not already classified as Big 4, Regional, or Specialist. The Specialists conducted approximately 15 percent of all A-133 audits in our sample, while the Big 4 and Regional firms each conducted 9 percent of the audits.[22] Organizations receiving at least $100,000 of public support are more likely to receive an audit by a Big 4 firm (15 percent of the Public Support Sample). In fact, untabulated results indicate that Big 4 firms audit the largest nonprofit organizations receiving federal funds. Using the Full Sample, the mean total assets of a Big 4 auditee is $405.9 million, while the mean total assets of a Regional auditee is $29.5 million. Regional firms, in turn, audit larger organizations than Specialists, which, in turn, audit larger organizations than the other accounting firms.

Table 3, Panel A provides descriptive statistics for the three samples. For the Full Sample, the mean (median) *SIZE* is $45.8 million ($2.4 million). The Public Support Sample is substantially larger, with a mean (median) *SIZE* of $101.6 million ($5.1 million). Across all three samples, organizations are relatively mature—the mean age for the Full Sample is 25 years. The median *INDIRECT SUPPORT* is $0 for all samples, indicating that many organizations do not receive any indirect support from federated fundraising campaigns. Note all of the continuous variables except *AGE* and *GROWTH* are right-skewed. Thus, we use natural log transformations for these variables in our analysis.

During our sample period, the A-133 audit provided four indicators of an internal control problem: (1) if the organization discloses a reportable condition related to financial reporting (*RC_FS*); (2) if any reportable condition related to financial reporting constitutes a material weakness (*MW_FS*); (3) if the organization discloses a reportable condition related to compliance with federal program requirements (*RC_GOV*); and (4) if any reportable condition related to federal program compliance constitutes a material weakness (*MW_GOV*). Material weaknesses are a subset of reportable conditions, representing the more severe internal control problems. In addition to theses four indicators, we also create a variable, *RC_ANY* (*MW_ANY*), which denotes the existence of a reportable condition (material weakness) over either financial reporting or federal program compliance.

Table 3, Panel B reports the frequency of each type of internal control problem. Given the relatively small number of observations with a material weakness, we focus primarily on reportable conditions. Overall, 14.86 percent of the sample discloses a reportable condition over financial reporting and 13.76 percent discloses a reportable condition of federal program compliance. Not surprisingly, there is significant overlap among organizations disclosing a financial statement internal control problem and organizations disclosing a federal program internal control problem. For comparison purposes, 14–15 percent of for-profit companies report a material weakness (Doyle et al. 2007a, Table 5).

V. RESULTS

Determinants of Internal Control Deficiencies

Table 4 reports simple correlations between our measures of internal control problems and organizational characteristics. As predicted, a going-concern paragraph in the audit opinion is positively associated with internal control deficiencies, while reporting a surplus is negatively associated with internal control deficiencies. Disclosure of internal control problems is positively associated with *GROWTH*, *RISK*, and *NEWGRANTEE* as expected. *SIZE* is negatively associated

[22] These frequencies are quite different from the frequencies in the for-profit sector. For example, Ashbaugh-Skaife et al. (2007) report that the six dominant audit suppliers account for 84.7 percent of the audits of public companies.

TABLE 3
Descriptive Statistics

Panel A: Descriptive Statistics for Continuous Variables

	n	Mean	Q1	Median	Q3
Full Sample					
PUBLIC SUPPORT	111,939	2,559,273	0	78,972	642,824
GOV CONTRIBUTIONS	111,939	4,094,543	0	765,158	2,444,370
DIRECT SUPPORT	111,939	2,278,322	0	39,397	466,141
INDIRECT SUPPORT	111,939	280,953	0	0	12,628
SIZE (TOTAL ASSETS)	127,988	45,783,365	861,596	2,354,736	7,124,068
AGE	125,345	25.13	12.00	22.00	34.00
GROWTH	127,306	1.11	0.97	1.02	1.14
PROGRAM REVENUE	127,988	16,409,849	43,038	380,984	2,749,254
FUNDRAISING EXP	111,939	265,676	0	0	47,164
PRICE	110,836	1.45	1.07	1.14	1.23
COMPLEXITY	127,998	2.19	2.00	2.00	3.00
Public Support Sample					
PUBLIC SUPPORT	47,318	5,482,325	252,395	705,244	2,532,311
GOV CONTRIBUTIONS	47,318	6,039,142	422,783	1,289,011	3,539,211
DIRECT SUPPORT	47,318	4,896,227	153,939	509,695	2,068,485
INDIRECT SUPPORT	47,318	586,099	0	0	140,223
SIZE (TOTAL ASSETS)	47,318	101,562,496	1,696,555	5,060,441	24,603,302
AGE	46,683	32.52	18.00	29.00	47.00
GROWTH	47,283	1.12	0.98	1.05	1.16
PROGRAM REVENUE	47,318	32,651,096	86814	1,118,323	9,882,294
FUNDRAISING EXP	47,318	574,088	0	50,235	290,720
PRICE	47,163	1.40	1.10	1.16	1.24
COMPLEXITY	47,318	2.72	2.00	3.00	3.00

(continued on next page)

Panel A: Descriptive Statistics for Continuous Variables

	n	Mean	Q1	Median	Q3
Government Sample					
PUBLIC SUPPORT	65,415	3,499,496	10,519	177,774	999,367
GOV CONTRIBUTIONS	65,415	6,245,366	698,919	1,549,013	3,994,750
DIRECT SUPPORT	65,415	3,163,269	2,620	106,665	755,560
INDIRECT SUPPORT	65,415	336,228	0	0	41,745
SIZE (TOTAL ASSETS)	65,415	66,015,301	890,679	2,681,559	9,182,541
AGE	64,343	27.16	14.00	24.00	35.00
GROWTH	65,389	1.12	0.97	1.04	1.17
PROGRAM REVENUE	65,415	19,277,339	20,000	314,772	2,874,398
FUNDRAISING EXP	65,415	368,364	0	0	93,063
PRICE	65,150	1.33	1.08	1.14	1.23
COMPLEXITY	65,415	2.62	2.00	3.00	3.00

Panel B: Descriptive Statistics for Indicator Variables

	Full Sample	Public Support Sample	Government Sample
RC_FS	14.86%	13.78%	14.30%
RC_ANY	18.91%	18.01%	17.99%
RC_GOV	13.76%	13.05%	13.34%
MW_FS	5.25%	4.35%	4.94%
MW_ANY	5.97%	5.10%	5.61%
MW_GOV	3.86%	3.28%	3.74%
GOINGCONCERNRISK	1.16%	0.82%	0.99%
SURPLUS	58.00%	67.93%	64.97%
RISK	34.74%	30.68%	32.33%
NEWGRANTEE	9.37%	6.74%	5.79%
LOBBYING	2.80%	4.80%	3.69%

All variables are defined in Table 1.

TABLE 4

Pearson Correlation Matrix

	RC_GOV	*COMPLEX*	*GOING CONCERN*	*SURPLUS*	*AGE*	*SIZE*	*GROWTH*
RC_FS	0.537 (<0.0001)	−0.015 (<0.0001)	0.083 (<0.0001)	−0.008 (0.0063)	0.007 (0.0099)	−0.041 (<0.0001)	0.012 (<0.0001)
RC_ANY	0.799 (<0.0001)	−0.019 (<0.0001)	0.080 (<0.0001)	−0.012 (<0.0001)	0.024 (<0.0001)	0.001 (0.7876)	0.010 (<0.0001)
RC_GOV		−0.012 (<0.0001)	0.068 (<0.0001)	−0.018 (<0.0001)	0.001 (0.7949)	−0.002 (0.4835)	0.007 (0.0225)
COMPLEXITY			−0.015 (<0.0001)	0.147 (<0.0001)	0.195 (<0.0001)	0.202 (<0.0001)	0.0111 (<0.0001)
GOINGCONCERNRISK				−0.059 (<0.0001)	−0.008 (0.0078)	−0.025 (<0.0001)	−0.0065 (0.0209)
SURPLUS					0.135 (<0.0001)	0.127 (<0.0001)	0.083 (<0.0001)
AGE						0.481 (<0.0001)	−0.022 (<0.0001)
SIZE							−0.015 (<0.0001)

	RISK	*NEW GRANTEE*	*PRICE*	*GDP*	*LOBBY*	*PUBLIC SUP*	*GOV CONT*
RC_FS	0.228 (<0.0001)	0.048 (<0.0001)	0.022 (<0.0001)	0.073 (<0.0001)	−0.015 (<0.0001)	−0.028 (<0.0001)	−0.010 (0.0079)
RC_ANY	0.248 (<0.0001)	0.042 (<0.0001)	0.023 (<0.0001)	0.102 (<0.0001)	−0.011 (<0.0001)	0.011 (0.0181)	−0.006 (0.1349)
RC_GOV	0.245 (<0.0001)	0.043 (<0.0001)	0.015 (<0.0001)	0.036 (<0.0001)	−0.011 (0.0007)	0.019 (0.0002)	−0.00002 (0.9960)
COMPLEXITY	0.003 (0.2354)	0.103 (<0.0001)	0.022 (<0.0001)	−0.232 (<0.0001)	0.059 (<0.0001)	0.182 (<0.0001)	0.176 (<0.0001)

(continued on next page)

	RISK	*NEW GRANTEE*	*PRICE*	*GDP*	*LOBBY*	*PUBLIC SUP*	*GOV CONT*
GOINGCONCERNRISK	0.056	0.003	0.015	−0.001	−0.015	−0.044	−0.012
	(<0.0001)	(0.2345)	(<0.0001)	(0.7380)	(<0.0001)	(<0.0001)	(0.0015)
SURPLUS	−0.025	0.014	0.026	0.037	0.047	0.131	0.086
	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)
AGE	−0.105	−0.153	−0.007	0.128	0.067	0.317	0.104
	(<0.0001)	(<0.0001)	(0.0343)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)
SIZE	−0.096	−0.143	0.033	0.108	0.131	0.548	0.197
	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)
GROWTH	0.021	0.0509	0.033	−0.005	0.0026	0.006	0.0008
	(<0.0001)	(<0.0001)	(<0.0001)	(0.0925)	(0.3529)	(0.1757)	(0.8355)
RISK		0.234	0.034	−0.051	−0.034	−0.081	−0.065
		(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)
NEWGRANTEE			0.079	−0.101	−0.026	−0.080	−0.107
			(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)
PRICE				0.006	0.009	0.008	−0.053
				(0.0587)	(0.0044)	(0.0862)	(<0.0001)
GDP					0.038	0.089	0.087
					(<0.0001)	(<0.0001)	(<0.0001)
LOBBY						0.119	0.060
						(<0.0001)	(<0.0001)

All variables are defined in Table 1. We use log form for all continuous variables except *AGE* and *GROWTH*. All variables except *PUBLIC SUPPORT* and *GOV CONTRIBUTIONS* use the full sample (n = 127,988). *PUBLIC SUPPORT* uses the public support sample (n = 47,318) and *GOV CONTRIBUTIONS* uses the government sample (n = 65,415). *RC_GOV* data is available for years beginning in 2001. p-values are in parentheses.

with *RC_FS*. The correlation between *COMPLEXITY* and the reportable condition measures is unexpectedly negative, which could be because *COMPLEXITY* is correlated with *SIZE*. For each of the empirical models estimated in the subsequent tables, we calculate variance inflation factors using ordinary least-squares and determine that multicollinearity is not a significant concern.

Table 5 presents the results from the first stage of our analysis. In Panel A, we estimate the probability of disclosing a reportable condition over financial reporting (*RC_FS*) as a function of organizational characteristics and audit detection factors, using a probit model for the three samples. The coefficients on *COMPLEXITY*, financial health (*GOINGCONCERNRISK* and *SURPLUS*), *SIZE*, *GROWTH*, and *RISK* have the predicted signs and are statistically significant. The coefficient on *NEWGRANTEE* is unexpectedly negative and significant. This negative coefficient results from the high association between *NEWGRANTEE* and *RISK*. When *RISK* is removed from the model, the coefficient on *NEWGRANTEE* is significantly positive.

In Table 5, Panel B, we estimate the probability of disclosing a reportable condition over either financial reporting or federal program compliance (*RC_ANY*). The results are similar to those in Panel A, except for the coefficients on *COMPLEXITY* in the Government Sample and *SIZE* in the Public Support and Government Samples. Overall, our evidence is consistent with the notion that less financially healthy and growing organizations disclose more internal control deficiencies. In addition, we provide some evidence that smaller and more complex organizations disclose more internal control problems.

The model in Table 5 also includes indicator variables for the type of auditor performing the A-133 audit. The coefficient on Big 4 auditors is reliably negative and significant, consistent with Keating et al. (2005). This result indicates that the probability of disclosing an internal control problem decreases if a Big 4 firm is used, and suggests that these audit firms selectively contract with certain high-quality nonprofit organizations. However, the coefficient on Regional firms is significantly positive, which indicates the likelihood of disclosing an internal control problem increases if a Regional audit firm is used.

We also estimate the probability of disclosing a reportable condition over federal program compliance (*RC_GOV*). The untabulated results are generally consistent with those in Table 5, with some minor exceptions. For the Full Sample, the coefficient on *COMPLEXITY* is not significant and for the Public Support Sample, the coefficients on *SIZE* and *GROWTH* are not significant. In addition, we estimate the probability of disclosing a material weakness (*MW_FS* and *MW_ANY*). Again, the results are generally consistent with those in Table 5. We find that all coefficients are statistically significant with the predicted sign, except for the coefficient on *GROWTH*, which is positive but not significant. Finally, we include the lagged internal control problem (either RC_FS_{t-1} or RC_ANY_{t-1}) and exclude *NEWGRANTEE* (because new grantees do not have prior-year internal control data). Not surprisingly, the coefficient on the lagged internal control problems is significantly positive. That is, if an organization reported an internal control weakness in the prior year, then the organization is more likely to report an internal control weakness in the current year.

Effect of Internal Control Deficiencies on Public Support

Table 6, Panel A reports the results from the second stage of our analysis. We estimate a regression of public support on the disclosure of reportable condition over financial reporting in the prior year and include the Inverse Mills ratio computed using the parameters from Table 5 for the Public Support Sample. For this table and all subsequent tables, we use Huber-White robust standard errors, where errors are clustered by organization.

In the first column ("Base") of Table 6, Panel A, we estimate the traditional Weisbrod and Dominguez (1986) model. Consistent with prior research, the coefficients on *FUNDRAISING EXP* and *AGE* are significantly positive, while the coefficient on *PRICE* is significantly negative. In

TABLE 5

Determinants of Internal Control Deficiencies

Panel A: Financial Statement Reportable Conditions

Variable	Pred. Sign	Full Sample	Public Support Sample	Government Sample
Intercept		−1.080***	−1.086***	−0.968***
		(398.828)	(117.427)	(161.340)
COMPLEXITY	+	0.081***	0.045***	0.029**
		(163.447)	(7.455)	(5.992)
GOINGCONCERNRISK	+	0.737***	0.882***	0.801***
		(455.994)	(173.643)	(238.627)
SURPLUS	−	−0.068***	−0.176***	−0.123***
		(48.240)	(111.267)	(78.667)
SIZE	−	−0.029***	−0.014***	−0.025***
		(91.230)	(6.670)	(34.136)
GROWTH	+	0.042***	0.052**	0.042***
		(15.460)	(5.930)	(7.260)
RISK	+	0.711***	0.726***	0.745***
		(5751.825)	(20960.732)	(3218.747)
NEWGRANTEE	+	−0.032**	−0.088***	−0.083***
		(4.270)	(8.624)	(9.810)
BIG4	?	−0.382***	−0.367***	−0.349***
		(316.585)	(150.812)	(142.823)
REGIONAL	?	0.280***	0.290***	0.243***
		(346.814)	(152.181)	(132.915)
SPECIALIST	?	−0.145***	−0.009	−0.102***
		(109.964)	(0.172)	(25.931)
Industry Indicators		Included	Included	Included
Year Indicators		Included	Included	Included
No. of Observations Used		127,236	47,281	65,364
No. of Observations with Internal Control Deficiencies		18,859	6,515	9,339
Likelihood Ratio		10356.535	3506.548	4857.398
		<0.0001	<0.0001	<0.0001
Percent Concordant		71.90%	71.30%	70.90%

Panel B: Any Reportable Conditions

Variable	Pred. Sign	Full Sample	Public Support Sample	Government Sample
Intercept		−1.385***	−1.562***	−1.334***
		(733.490)	(275.569)	(340.810)
COMPLEXITY	+	0.046***	0.046***	−0.010
		(58.822)	(8.833)	(0.749)
GOINGCONCERNRISK	+	0.705***	0.815***	0.786***
		(424.905)	(149.436)	(232.546)
SURPLUS	−	−0.081***	−0.148***	−0.120***
		(75.011)	(87.255)	(83.362)
SIZE	−	−0.008***	0.010*	−0.002
		(8.672)	(3.821)	(0.338)

(continued on next page)

Panel B: Any Reportable Conditions

Variable	Pred. Sign	Full Sample	Public Support Sample	Government Sample
GROWTH	+	0.046	0.061***	0.048***
		(19.673)	(8.864)	(10.339)
RISK	+	0.752***	0.753***	0.771***
		(7243.628)	(2531.856)	(3820.811)
NEWGRANTEE	+	−0.032*	−0.049*	−0.080***
		(4.668)	(2.893)	(9.469)
BIG4	?	−0.070***	−0.103***	−0.089***
		(15.905)	(15.623)	(12.412)
REGIONAL	?	0.288***	0.308***	0.251***
		(404.169)	(192.368)	(155.280)
SPECIALIST	?	−0.083***	−0.007	−0.049***
		(40.297)	(0.096)	(6.993)
Industry Indicators		Included	Included	Included
Year Indicators		Included	Included	Included
No. of Observations Used		127,236	47,281	65,364
No. of Observations with Internal Control Deficiencies		23,996	8,516	11,749
Likelihood Ratio		11345.371	3954.150	5346.047
		<0.0001	<0.0001	<0.0001
Percent Concordant		70.70%	70.30%	69.90%

*, **, *** Indicates statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

All variables are defined in Table 1. We use log form for all continuous variables except *GROWTH*. Wald Chi-squared statistics are reported in parentheses.

model (1), we include an indicator variable for the existence of a reportable condition over financial reporting. The coefficient on RC_FS_{t-1} (−0.210) is significantly negative. In model (2), we add *GOV CONTRIBUTIONS*, *PROGRAM REVENUE*, and $PUBLIC\ SUPPORT_{t-1}$ which are all significantly associated with $PUBLIC\ SUPPORT_t$. Nevertheless, the coefficient on RC_FS_{t-1} is still significantly negative. Note that by including lagged public support, we are essentially estimating the *change* in public support associated with disclosing an internal control problem. Our evidence suggests that, all else equal, reporting internal control problems over financial reporting is associated with 3.8 percent less public support on average. In model (3) of Panel A, we estimate the influence of reporting the more severe material weakness over financial reporting on subsequent public support. Again, the coefficient is significantly negative (−0.056). As indicated in Table 6, Panel B, we find similar results when we use *RC_ANY*. The evidence in Table 6 is consistent with our hypothesis that reportable conditions related to the financial statements have a detrimental effect on subsequent support.

Next, we examine the components of *PUBLIC SUPPORT* in models (4) and (5). Both Panels A and B of Table 6 suggest that that *DIRECT SUPPORT* is negatively associated with the existence of an internal control weaknesses. The evidence for *INDIRECT SUPPORT* is mixed; Panel A reports a significantly negative coefficient, while Panel B reports an insignificant coefficient. Interestingly, the frequency of internal control problems is not statistically different for firms with

TABLE 6

The Effect of Internal Control Deficiencies on Public Support

Panel A: Financial Statement Internal Control Weaknesses

Variable	*PUBLIC SUPPORT* Base	*PUBLIC SUPPORT* (1)	*PUBLIC SUPPORT* (2)	*PUBLIC SUPPORT* (3)	*DIRECT* (4)	*INDIRECT* (5)
Intercept	13.079***	13.172***	0.699***	0.692***	0.511***	0.048
	(132.60)	(1212.15)	(13.91)	(13.80)	(10.98)	(1.63)
RC_FS_{t-1}		−0.0210***	−0.038***		−0.034**	−0.028**
		(−8.07)	(−3.09)		(−2.33)	(−2.14)
MW_FS_{t-1}				−0.056***		
				(−2.58)		
$FUNDRAISING\ EXP_{t-1}$	0.134***	0.135***	0.022***	0.022***	0.026***	0.003***
	(39.10)	(37.17)	(22.49)	(22.54)	(21.66)	(3.78)
$PRICE_{t-1}$	−0.151***	−0.226***	−0.058	−0.059	−0.031	−0.016
	(−4.72)	(−4.30)	(−1.46)	(−1.47)	(−0.73)	(−1.41)
AGE	0.016***	0.016***	0.002***	0.002***	0.001***	0.0004**
	(18.88)	(17.42)	(10.20)	(10.21)	(6.56)	(2.00)
$GOV\ CONTRIBUTIONS_{t-1}$			0.009***	0.009***	0.009***	0.002**
			(9.74)	(9.76)	(8.57)	(1.95)
$PROGRAM\ REVENUE_{t-1}$			0.010***	0.009***	0.006***	0.001
			(8.91)	(9.76)	(5.23)	(0.91)
$PUBLIC\ SUPPORT_{t-1}$			0.922***	0.922***		
			(266.21)	(266.20)		
$DIRECT\ SUPPORT_{t-1}$					0.933***	
					(309.96)	
$INDIRECT\ SUPPORT_{t-1}$						0.985***
						(1168.79)
Inverse Mills		0.064***	0.006	0.008	0.010	−0.006
		(4.87)	(0.91)	(0.58)	(1.15)	(−0.83)
Industry Indicators	Included	Included	Included	Included	Included	Included
Year Indicators	Included	Included	Included	Included	Included	Included

(continued on next page)

Panel A: Financial Statement Internal Control Weaknesses

	PUBLIC SUPPORT				*DIRECT*	*INDIRECT*
Variable	**Base**	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
Number of Observations	44,353	40,030	37,878	37,878	37,588	36,601
R^2	35.59%	36.40%	83.20%	83.19%	90.13%	97.77%

Panel B: Any Internal Control Weaknesses

	PUBLIC SUPPORT				*DIRECT*	*INDIRECT*
Variable	**Base**	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
Intercept	13.079***	13.146***	0.693***	0.690***	0.509***	0.038
	(132.60)	(121.80)	(13.83)	(13.77)	(10.97)	(1.31)
RC_ANY_{t-1}		−0.086***	−0.028**		−0.027**	−0.018
		(−3.77)	(−2.50)		(−2.17)	(−1.50)
MW_ANY_{t-1}				−0.046**		
				(−2.29)		
FUNDRAISING EXP_{t-1}	0.134***	0.136***	0.022***	0.022***	0.026***	0.003***
	(39.10)	(37.26)	(22.51)	(22.54)	(21.65)	(3.93)
$PRICE_{t-1}$	−0.151***	−0.227***	−0.058	−0.059	−0.030	−0.016
	(−4.72)	(−4.25)	(−1.46)	(−1.46)	(−0.72)	(−1.38)
AGE	0.016***	0.016***	0.002***	0.002***	0.001***	0.0005**
	(18.88)	(17.47)	(10.20)	(10.21)	(6.59)	(2.10)
GOV $CONTRIBUTIONS_{t-1}$			0.009***	0.009***	0.009***	0.002**
			(9.69)	(9.74)	(8.51)	(2.03)
PROGRAM $REVENUE_{t-1}$			0.010***	0.010***	0.006***	0.001
			(8.97)	(8.94)	(5.25)	(0.85)
PUBLIC $SUPPORT_{t-1}$			0.922***	0.922***		
			(266.52)	(266.27)		
DIRECT $SUPPORT_{t-1}$					0.934***	
					(310.24)	

(continued on next page)

Panel B: Any Internal Control Weaknesses

		PUBLIC SUPPORT			*DIRECT*	*INDIRECT*
Variable	**Base**	(1)	(2)	(3)	(4)	(5)
INDIRECT SUPPORT$_{t-1}$						0.985***
						(1171.170)
Inverse Mills		0.036***	0.008	0.008	0.009	−0.005
		(2.97)	(1.23)	(0.83)	(1.23)	(−0.80)
Industry Indicators	Included	Included	Included	Included	Included	Included
Year Indicators	Included	Included	Included	Included	Included	Included
Number of Observations	44,353	40,028	37,877	37,878	37,587	36,600
R^2	35.59%	36.32%	83.21%	83.19%	90.14%	97.77%

*, **, *** Indicates statistical significance at the 0.10, 0.05, or 0.01 level, respectively. All variables are defined in Table 1. We use log form for all continuous variables except *AGE*. Influential observations, identified as studentized residuals greater than 3, are removed. t-statistics are reported in parentheses. We use Huber-White robust standard errors clustered by organization.

direct support compared to firms with indirect support.[23] Those organizing and contributing to federated fund-raising campaigns do not appear to consider internal control problems as part of the giving decision in a more sophisticated manner than other donors.

We also estimate the effect of disclosing a reportable condition over federal program compliance on subsequent support (untabulated). The coefficient on *RC_GOV* is negative but not significant ($p = 0.16$). This result suggests that giving decisions made by individuals, corporations, and foundations are highly associated with internal control problems over financial reporting but not with internal control problems over federal program compliance.

We conduct a series of robustness tests on model (2) in Table 6, Panel A and Panel B. We control for macroeconomic factors by including GDP and state indicator variables. We address the possibility that low financial reserves are associated with both public support and the existence of an internal control problem by including *GOINGCONCERNRISK*. We include *SIZE*, as well as scale all dollar-denominated continuous variables by total revenue. Finally, we change the sample selection criteria to include all organizations with public support over $1,000, instead of $100,000. For each of these tests, we continue to find negative and significant coefficients on *RC_FS* and *RC_ANY*, suggesting donors directly or indirectly use internal control information in their giving decisions.

Finally, we investigate the association between internal control problems and subsequent public support across the six NTEE industries listed in Table 1. We find negative and significant coefficients on both *RC_FS* and *RC_ANY* for Education and Health. The coefficients on internal controls problems for the remaining industries are insignificant. According to Wing et al. (2008), excluding Religion,[24] Education receives the largest percentage of total public support, followed next by Health. Thus, it appears that internal control problems influence industries that receive substantial contributions from donors.

Effect of Internal Control Deficiencies on Government Contributions

Table 7 reports the results from estimating a regression of government contributions on *RC_FS*, *RC_ANY*, and *RC_GOV* for the Government Sample. We again include the inverse Mills ratio computed using the parameters from Table 5. The coefficients on *RC_FS* (−0.017) and *RC_ANY* (−0.017) are negative and significant. Unlike with public support, the coefficient on *RC_GOV* (−0.021) is also significantly negative. These results suggest that government agencies do use information regarding internal controls over federal program compliance to make funding decisions. When we use material weaknesses instead of reportable conditions, only the coefficient on *MW_GOV* is significant.

In addition, the coefficient on *FUNDRAISING EXP* is significantly negative, consistent with Andreoni and Payne (2003). The coefficient on *LOBBYING* is positive, suggesting that organizations that engage in activities to influence legislation receive more government grants. The coefficient on GDP is negative. Overall, government contributions to nonprofit organizations appear to be a function of political and economic factors, as well as the organization's perceived ability to fulfill its mission as signaled by the results of its internal control audit.

The results in Table 7 are consistent using alternative specifications of Equation (3). In

[23] In fact, 20.7 percent of organizations that receive only indirect support disclose an internal control problem, while 19.1 percent of organizations that receive only direct support disclose a problem. Of the organizations that receive both direct and indirect support, 16.8 percent disclose an internal control problem. These percentages suggest that any differential effects of internal controls problems on direct and indirect support are not caused by superior selection methods of federated fundraising campaigns (i.e., these campaigns do not select out of giving to organizations with problems).

[24] Religion receives the largest amount of public support. Most religious organizations are not included in our sample because religious institutions are exempt from the Form 990 filing requirement and most do not receive federal funding.

TABLE 7

The Effect of Internal Control Deficiencies on Government Contributions

Variable	*GOV CONTRIBUTIONS*					
Intercept	0.561***	0.558***	0.587***	0.556***	0.556***	0.588***
	(10.39)	(10.33)	(12.45)	(10.30)	(10.30)	(10.25)
RC_FS_{t-1}	−0.017**					
	(−2.42)					
RC_ANY_{t-1}		−0.017***				
		(−2.61)				
RC_GOV_{t-1}			−0.021**			
			(−2.52)			
MW_FS_{t-1}				−0.013		
				(−1.18)		
MW_ANY_{t-1}					−0.013	
					(−1.19)	
MW_GOV_{t-1}						−0.028*
						(−1.86)
$FUNDRAISING\ EXP_{t-1}$	−0.001*	−0.001*	−0.0005	−0.001*	−0.001*	−0.001
	(−1.90)	(−1.91)	(−0.90)	(−1.88)	(−1.88)	(−0.89)
$PRICE_{t-1}$	−0.026*	−0.025*	−0.024	−0.026*	−0.026*	−0.024
	(−1.76)	(−1.75)	(−1.46)	(−1.76)	(−1.77)	(−1.46)
$LOBBYING_t$	0.018**	0.019**	0.020*	0.018**	0.018**	0.019*
	(2.07)	(2.06)	(1.83)	−2.04	−2.05	−2.05
AGE	0.000	0.000	0.000	0.000	0.000	0.000
	(−0.16)	(−0.14)	(−0.43)	(−0.12)	(−0.13)	(−0.43)
$PUBLIC\ SUPPORT_{t-1}$	0.005***	0.005***	0.004***	0.005***	0.005***	0.004***
	(8.72)	(8.73)	(6.29)	(8.74)	(8.74)	(6.28)
$PROGRAM\ REVENUE_{t-1}$	−0.002***	−0.002***	−0.002***	−0.002***	−0.002***	−0.002***
	(−3.66)	(−3.64)	(−2.73)	(−3.63)	(−3.62)	(−2.72)
$GOV\ CONTRIBUTIONS_{t-1}$	0.978***	0.978***	0.977***	0.978***	0.978***	0.976***
	(518.71)	(518.82)	(454.62)	(518.72)	(518.36)	(554.35)

(continued on next page)

TABLE 7 (continued)

Variable	*GOV CONTRIBUTIONS*					
GDP	−0.014***	−0.013***	−0.016***	−0.014***	−0.014***	−0.016***
	(−9.34)	(−9.19)	(−8.29)	(−9.28)	(−9.26)	(−8.47)
Inverse Mills	0.013***	0.011***	0.012***	0.008	0.010*	0.015**
	(3.03)	(3.00)	(2.98)	(1.50)	(1.82)	(2.20)
State Indicators	Included	Included	Included	Included	Included	Included
Industry Indicators	Included	Included	Included	Included	Included	Included
Number of Observations	52,411	52,411	40,369	52,411	52,411	40,369
R^2	89.73%	89.73%	89.71%	89.73%	89.73%	89.71%

*, **, *** Indicates statistical significance at the 0.10, 0.05, or 0.01 level, respectively. All variables are defined in Table 1. We use log form for all continuous variables except *AGE*. Influential observations, identified as studentized residuals greater than three, are removed. t-statistics are reported in parentheses. We use Huber-White robust standard errors clustered by organization.

particular, we include *GOINGCONCERNRISK* and *SIZE*, scale the continuous variables by total revenue, and change the sample selection criteria to include all organizations with government contributions over $1,000 instead of $100,000. For each of these tests, we continue to find negative and significant coefficients on the reportable condition indicators, suggesting that government agencies use internal control information in their funding decisions.

In untabulated results, we next investigate the link between internal control problems and subsequent government contributions for the five most frequent oversight agencies in the Government Sample.[25] The most frequent oversight agencies (percentage of sample) are the Department of Health and Human Services (46 percent), the Department of Housing and Urban Development (16 percent), the Department of Education (14 percent), the Department of Agriculture (5 percent) and the Department of Justice (3 percent). No other federal agency represents more than 1 percent of the sample. The coefficients on all three measures of reportable conditions are significantly negative for organizations overseen by HUD. In addition, the coefficient on reportable conditions over financial reporting is significantly negative for organizations overseen by the Department of Agriculture. The coefficients on reportable conditions for organizations overseen by HHS, by far the most frequent oversight agency, are negative but not significant.[26] The coefficients on reportable coefficients for the remaining most frequent agencies are also insignificant.

VI. CONCLUSION

This study examines the causes and consequences of internal control weaknesses in nonprofit organizations. The nonprofit sector provides a useful setting to examine the effects of internal control disclosures because charitable giving by donors provides direct evidence of stakeholder reactions to such disclosures. We first document that the likelihood of reporting an internal control problem increases for nonprofit organizations that are complex, growing, smaller, and in poor financial health. We then present evidence that weak internal controls over financial reporting are negatively associated with subsequent public support and government contributions received after controlling for the current level of support and other factors influencing contributions. Thus, internal control information appears to affect, either directly or indirectly, the funders' giving decisions. Internal control reporting by nonprofits has been required for two decades. Our results are generalizable to the for-profit sector because they provide long-term evidence that stakeholders do indeed use internal control information to evaluate organizations.

Our specific results may interest several constituencies. First, the IRS and other regulators are reformulating laws in an attempt to increase public confidence in the integrity of exempt organizations. Second, donors want to make more informed charitable decisions. Third, watchdog groups promulgate standards to evaluate an organization's effectiveness in achieving its mission. These standards may encourage organizations to underinvest in infrastructure in the short-run. In fact, Hager et al. (2004) argue that pressure from donors and watchdog groups to maximize mission-related spending and limit overhead costs to artificially low levels is detrimental in the long-run.

[25] We define the oversight agency as the cognizant agency if a cognizant agency exists; otherwise it is the oversight agency designated in the A-133 report. The oversight agency is likely, but not necessarily, the predominant supplier of government contributions. To the extent that a nonprofit organization is complex and receives contributions from several federal agencies or receives most federal funding from contract revenue rather than from contributions, the oversight agency will be less meaningful for purposes of this test.

[26] Recall that Medicare and Medicaid are program service revenue and not part of our tests. One possible explanation for the lack of significant results for HHS is that the majority of HHS discretionary grant dollars are allocated to scientific research (http://www.hhs.gov/grants/). These are generally multiperiod grants. Thus, our tests may not properly cover the decision window for these grants.

Our evidence is consistent with the notion that short-term savings on administrative expenses (i.e., establishing internal controls) can ultimately have negative consequences on the organization's donor support and, thus, the organization's ability to deliver services.

Finally, nonprofit managers and board members need to understand the risks of failing to meet donors' and government agencies' expectations with regards to accountability. We estimate that, all else equal, organizations with internal control problems receive 3.8 percent less public support and 2.1 percent less government support. According to the National Council on Nonprofits (2009), the estimated cost of an audit for an organization with revenue of $600,000 is $12,000–$20,000. Audit costs likely increase in the presence of internal control problems (Hogan and Wilkins 2008). Thus, in some cases, a cost-benefit analysis indicates an overall benefit from conducting periodic, thorough internal reviews of internal controls. If attestations of internal controls by external auditors are cost-prohibitive, then nonprofit organizations can seek in-kind support to help them improve their internal controls. For example, technology companies often donate technical support to nonprofit organizations. Similarly, other corporate donors with Sarbanes-Oxley experience can provide guidance on creating and maintaining adequate internal control systems.

Internal control in the nonprofit sector is a relatively unexplored area for researchers and there are many questions left to be addressed. We show that the disclosure of any internal control weakness is associated with future declines in public support, but leave open the question of which types of donors (i.e., foundations, corporations, or individuals) respond to the internal control information. Likewise, a more refined analysis would assist in determining which federal, state, and local agencies use the A-133 data. Finally, further research is needed to investigate how internal control weaknesses influence other aspects of a nonprofit organization's operations, including earnings management and executive compensation.

REFERENCES

American Institute of Certified Public Accountants (AICPA). 2003. *Audits of States, Local Governments and Not-For-Profit Organizations Receiving Federal Awards.* New York, NY: American Institute of Certified Public Accountants.

———. 2009. *Government Auditing Standards and Circular A-133 Audits.* New York, NY: American Institute of Certified Public Accountants.

———. 2010. *AICPA Professional Standards.* New York, NY: American Institute of Certified Public Accountants.

Andreoni, J. 1990. Impure altruism and donations to public goods: A theory of warm-glow giving. *The Economic Journal* 100 (401): 464–477.

———, and A. Payne. 2003. Do government grants to private charities crowd out giving or fund-raising? *The American Economic Review* 93 (3): 792–812.

Ashbaugh-Skaife, H., D. Collins, and W. Kinney. 2007. The discovery and reporting of internal control deficiencies prior to SOX-mandated audits. *Journal of Accounting and Economics* 44 (1–2): 166–192.

———, ———, ———, and R. LaFond. 2009. The effect of SOX internal control deficiencies on firm risk and cost of equity. *Journal of Accounting Research* 47 (1): 1–43.

Atkinson, R. 1998. Unsettled standing: Who (else) should enforce the duties of charitable fiduciaries? *The Journal of Corporation Law* 23 (4): 655–700.

Baber, W., A. Roberts, and G. Visvanathan. 2001. Charitable organizations' strategies and program-spending ratios. *Accounting Horizons* 15 (4): 329–343.

Beneish, M., M. Billings, and L. Hodder. 2008. Internal control weaknesses and information uncertainty. *The Accounting Review* 83 (3): 665–703.

Bowman, W. 2006. Should donors care about overhead costs? Do they care? *Nonprofit and Voluntary Sector Quarterly* 35 (2): 288–310.

Brown, K. 2007. Letter to Senator Charles E. Grassley from Acting Commissioner of IRS. Dated June 28, 2007. Available at: http://finance.senate.gov/press/Gpress/2007/prg072307a.pdf.

Dicker, F., and S. Goldenberg. 2009. Cuomo's pay-to-play rap for nonprofits. *New York Post* (October 23).

Doyle, J., W. Ge, and S. McVay. 2007a. Determinants of weaknesses in internal control over financial reporting. *Journal of Accounting and Economics* 44 (1–2): 193–223.

Fama, E., and M. Jensen. 1983. Separation of ownership and control. *The Journal of Law & Economics* 26 (2): 301–325.

Fishman, J. 2003. Improving charitable accountability. *Maryland Law Review* 62: 218–287.

Fremont-Smith, M. 2007. The search for greater accountability of nonprofit organizations: Recent legal developments and proposals for change. *Fordham Law Review* 76: 609–646.

Ge, W., and S. McVay. 2005. The disclosure of material weaknesses in internal control after the Sarbanes-Oxley Act. *Accounting Horizons* 19 (3): 137–158.

Grassley, C. 2006. Strengthening the nonprofit sector. *The Hill* (July 12): 26.

Greenlee, J., and K. Brown. 1999. The impact of accounting information on contributions to charitable organizations. *Research in Accounting Regulation* 13: 111–125.

———, M. Fischer, T. Gordon, and E. Keating. 2007. An investigation of fraud in nonprofit organizations: Occurrences and deterrents. *Nonprofit and Voluntary Sector Quarterly* 36 (4): 676–694.

Hager, M., T. Pollak, and K. Wing. 2004. Getting what we pay for: Low overhead limits nonprofit effectiveness. Nonprofit Overhead Cost Project Brief No. 3. Available at: http://www.urban.org/publications/311044.html.

Hammersley, J., L. Myers, and C. Shakespeare. 2008. Market reactions to the disclosure of internal control weaknesses and to the characteristics of those weaknesses under Section 302 of the Sarbanes Oxley Act of 2002. *Review of Accounting Studies* 13 (1): 141–165.

Hansmann, H. 1981. Reforming nonprofit corporation law. *University of Pennsylvania Law Review* 129: 497–623.

Heckman, J. 1979. Sample selection bias as a specification error. *Econometrica* 47 (1): 153–161.

Hogan, C., and M. Wilkins. 2008. Evidence on the audit risk model: Do auditors increase audit fees in the presence of internal control deficiencies. *Contemporary Accounting Research* 25 (1): 219–242.

Irvin, R. 2005. State regulation of nonprofit organizations: Accountability regardless of outcome. *Nonprofit and Voluntary Sector Quarterly* 34 (2): 161–178.

Iyer, V., and A. Watkins. 2008. Adoption of Sarbanes-Oxley measures by nonprofit organizations: An empirical study. *Accounting Horizons* 22 (3): 255–277.

Keating, E., M. Fischer, T. Gordon, and J. Greenlee. 2005. The Single Audit Act: How compliant are nonprofit organizations? *Accounting and Financial Management* 17 (3): 285–309.

Khanna, J., and T. Sandler. 2000. Partners in giving: The crowding-in effect of UK government grants. *European Economic Review* 44 (8): 1543–1556.

Khumawala, S., L. Parsons, and T. Gordon. 2005. Assessing the quality of not-for-profit efficiency ratios: Do donors use joint cost allocation disclosures? *Journal of Accounting, Auditing & Finance* 20 (3): 287–309.

Kitching, K. 2009. Audit value and charitable organizations. *Journal of Accounting and Public Policy* 28 (6): 510–524.

Mulligan, L. 2007. What's good for the goose is not good for the gander: Sarbanes-Oxley-style nonprofit reforms. *Michigan Law Review* 105 (8): 1981–2009.

National Council of Nonprofits. 2009. Special state focus: Audit thresholds. *Nonprofit Policy News* (April). Available at: http://www.councilofnonprofits.org/?q=policynews/april2009#state.

Ogneva, M., K. Subramanyam, and K. Raghunandan. 2007. Internal control weakness and cost of equity: Evidence from SOX section 404 disclosures. *The Accounting Review* 82 (5): 1255–1297.

Okten, C., and B. Weisbrod. 2000. Determinants of donations in private nonprofit markets. *Journal of Public Economics* 75 (2): 255–272.

Ostrower, F. 2007. *Nonprofit Governance in the United States: Findings on Performance and Accountability from the First National Representative Study*. Washington, D.C.: The Urban Institute.

Posnett, J., and T. Sandler. 1989. Demand for charity donations in private non-profit markets: The case of the U.K. *Journal of Public Economics* 40 (2): 187–200.

PricewaterhouseCoopers. 2006. Internal control readiness: Why you need to act today. Available at: http://www.pwc.com.

Reiser, D. 2005. Enron.org: Why Sarbanes-Oxley will not ensure comprehensive nonprofit accountability. *U.C. Davis Law Review* 38: 205–280.

Ribar, D., and M. Wilhelm. 2002. Altruistic and joy-of-giving motivations in charitable behavior. *The Journal of Political Economy* 110 (2): 425–457.

Tinkelman, D. 1998. Difference in sensitivity of financial statement users to joint cost allocations: The case of nonprofit organizations. *Journal of Accounting, Auditing & Finance* 13 (4): 377–394.

———. 2004. Using non-profit organization-level financial data to infer managers' fundraising strategies. *Journal of Public Economics* 88 (9–10): 2181–2192.

———, and K. Mankaney. 2007. When is administrative efficiency associated with charitable donations? *Nonprofit and Voluntary Sector Quarterly* 36 (1): 41–64.

Weisbrod, B., and N. Dominguez. 1986. Demand for collective goods in private nonprofit markets: Can fundraising expenditures help overcome free-rider behavior? *Journal of Public Economics* 30 (1): 83–95.

Wing, K., T. Pollak, and A. Blackwood. 2008. *The Nonprofit Almanac*. Washington, D.C.: Urban Institute Press.

Yetman, M., 2008. Are donors misled by low quality financial reports? Working paper, University of California, Davis.

THE ACCOUNTING REVIEW
Vol. 86, No. 1
2011
pp. 359–361

American Accounting Association
DOI: 10.2308/accr.00000025

BOOK REVIEWS

Stephen A. Zeff, Editor

Editor's note: Two copies of books for review should be sent to the Book Review Editor: Stephen A. Zeff, Rice University, Jesse H. Jones Graduate School of Business, 6100 Main St., Houston, TX 77005. The policy of *The Accounting Review* is to publish only those reviews solicited by the Book Review Editor. Unsolicited reviews will not be accepted.

MATTHEW GILL, *Accountants' Truth: Knowledge and Ethics in the Financial World* (Oxford, U.K.: Oxford University Press, 2009, ISBN 978-0-19-954714-2, pp. x, 198).

More than two decades ago, I closed a book on the professions with the remark that "surely accounting is today far more socially important than medicine." There were then few studies of the accounting profession and, above all, no serious theoretical analyses of accounting as the intellectual activity that both measured and normatively defined the economic phenomena at the center of modern life. In the years since, there have been a number of such studies and indeed a journal (*Accounting, Organizations and Society*) that has focused attention on such questions. Matthew Gill's *Accountants' Truth* falls in this new tradition. Because the book is written primarily for an audience in accounting rather than sociology, however, I am an awkward reviewer. But perhaps the accountant reader will find it interesting to know what strikes a sociologist about Gill's book.

The book describes the intellectual world of a very particular group: 20 accountants, snowball-sampled five each from the Big 4's London offices. They are all young men, nearly all ICAEW-qualified, nearly all within a few years of their original training. Since many potential interviewees declined to participate, the group is somewhat self-selected. Himself an ICAEW accountant, Gill interviewed these 20 in detail, including in his interview protocol a discussion of an accounting problem-scenario sent ahead. The book has six substantive chapters: a theoretical discussion of performance and truthfulness, an empirical analysis of reactions to the scenario, chapter-length discussions of technocratism, pragmatism, and professionalism, and a chapter on ethics.

Gill argues that accountants inevitably experience a tension between technocratism (the attempt to reduce accounting to mere rules, abstracting accounting from accountants) and pragmatism (the decision to bracket accounting's "truth" and to understand accounting practice through alternative metaphors such as strategy, sport, or family). The problem with rules is that rationalization in accounting simply makes accountants more and more alienated from the actual practice of accounting; rules make accountants automata, reducing professional commitment as well as ambiguity. As a result, "professionalism" becomes for these young men a kind of refuge—vacuous, but safe nonetheless—from the ethical problems that their alienated selves can see but lack the desire to rectify. "Ethics" becomes a matter of subtly shifting responsibility to the firm ("employee ethics") or onto the law ("legal ethics"). Beyond these lies equivocation, yet another technique for handling the alienation inevitable in accounting work. Yet in the end Gill argues that this state of affairs is quite understandable, given the actual situations of his respondents, and that plans to shift it by the usual means of further rationalization and regulation are self-defeating. He devotes a final chapter to potential fixes for alienation, following the late Eliot Freidson—longtime dean of the sociologists of professions—in calling for a renewed commitment to a deeper and more comprehensive professionalism.

Although I found the book very interesting, I worry that what may seem interesting to me—the demonstrations of exactly how it is that accounting works as a particular discourse and how accounting "facts" are constructed—may be everyday common sense to the accounting reader. Conversely, the more general claims—for example, the claim that knowledge consists of a best stab at representing something about which there can be no final truth—are common sense in sociology, although they may be new to the accountant reader. The book doesn't

aim to further develop such general claims, but rather to apply them in the particular world of accounting (that's why it seems to me a book for accountants). And although its central idea—that increasing rationalization paradoxically increases alienation among accountants and may even facilitate equivocation—is quite interesting, that claim cannot (at least for a sociologist) be empirically established or even fully explored on the basis of only 20 interviews.

Such a reaction is inevitable in the sociologist reader. Discourse and construction are old news to me, and I have never thought that contemporary accountants merely discover facts, although I do know that a century ago accountants on-site counted physical inventory. What I wanted in the book was five times as many interviews, quota-sampled across a variety of types of accounting, and a much more detailed analysis of how different aspects of the discourse and practice of accounting go together differently (or similarly, if that is the case) for different kinds of accountants. That is, I wanted much more data and the analysis a sociologist would have written.

But that's a narrow disciplinary worry. The advantage of a very small sample is that it allows respondents to become full and complex individuals. Yet Gill's narrative does not always take full advantage of that complexity, both because he takes his respondents as a group emblematic of accounting in the abstract and because the organization of the book around substantive themes (rather than the complexities of individuals' adaptations) hides those very particular complexities from us. As a result, having chosen respondents very closely grouped in age, gender, professional seniority, and employer, Gill leaves us wondering how far these results generalize. Perhaps more disturbing, he tells us early on that many of these young men will leave for other occupations in short order; accountancy is for them a life stage, a preparation for something else. But given that, how can we take their "discourse" as that of accountancy in general? Gill claims that by looking at the core of the recently trained elite he will find the cutting edge of accountants' discourse. True enough, but one cannot help thinking that these are very young men, interviewed by a young man, and that their love of the "sport" of tax evasion, their appreciation for the clever "wheeze," and their cavalier attitude toward their society and its non-economic institutions may derive from a juvenile masculinity that will wear off as life teaches them a few nasty lessons.

The sample raises another issue. It is to be sure a sample of accountants, but it is also a sample of people who work for large organizations. The refuge in technocracy and pragmatism, the protective armor of employee ethics and avoidance of responsibility—these are the stuff of bureaucratic life. Indeed, Gill often invokes Robert Jackall's (1988) brilliant *Moral Mazes*, which concerns considerably older middle and upper managers of large organizations. And like Jackall's managers, Gill's accountants will be able to "outrun" whatever problems they create, in their case by leaving the occupation entirely rather than by new appointments elsewhere. One worries, then, that Gill's actual topic may be organization men, not accountants in particular.

My insistence on surroundings of course reflects my own work about the professions. I wanted to hear much more about accountants' relations with lawyers, tax advisors, business consultants, and other competitors in this work area. The book is very good on relations with clients, but other than a few words about the law, it doesn't tell us a lot about the other professions crowding the accountants, nor about those (e.g., business consultants) whom the accountants would themselves like to crowd out (a battle that has been going on at least since the 1920s, at least in the U.S.).

As someone whose theories are rooted in the historical evolution of professions over time, I also wanted more background on the history of accounting and on the historical moment (after neoliberalism and the so-called Big Bang in English legal affairs) in which these young men found themselves. Indeed, the book's argument that rationalization increases alienation cannot, ultimately, be established on the basis of purely contemporary data. It requires data over time.

Finally, it is of the nature of a book written for an accounting audience that it does not ask what, for a sociologist, would be the major theoretical questions. But nonetheless an accountant reader might wonder what those sociological questions are. For me, they involve the second-order clients or, to put it in John Dewey's term, the "public" of accounting. (A Marxist would call it the political economy of accounting.) Although Gill starts us with some general words about principals and agents, he moves very quickly into a data analysis that takes the larger social situation of his accountants for granted. But public accounting started not because clients sought it, but because investors demanded it. In the early years of English public accounting, when 80 percent to 90 percent of new joint stock companies failed annually and vast amounts of British capital were being invested in projects overseas, investors demanded public accounting as a means of knowing something real about the objects of their investments. Early English accounting knew that its principal "client" was the investor, not the firm audited. Today, of course, the situation is quite different. It is not clear for which principal accountants are today the agents, nor indeed which among the investors—smart money, dumb money—is the "user" envisioned by accountants when they decide to locate this or that exceptional item above or below the line. And of course the Inland Revenue has also emerged as a crucial audience for reports, as is the legal system that holds the ultimate power.

This complex arena in which the various consumers of balance sheets transact and argue with one another is never seriously analyzed, by either the respondents or by Gill himself. The general social setting of accounting, how it is paid for, who reads it, what are its distributional consequences for the society at large: none of these is treated as problematic, perhaps because they are so commonsensical to the presumed audience. But to a sociologist, these matters are the main event. The principal force driving the complexities and difficulties that Gill

portrays so well is the simple fact that accountants are providing information on the basis of which some people will try to make money at the expense of others. The audited firms that provide them with information, the short-run looter-investors, the long-run disattentive investors, the tax officials aiming to fund a government, the workers wondering about pension funding: all these people are trying quite reasonably to acquire money that others think is more properly theirs, or theirs for the taking, or whatever. These various people have widely varying levels of expertise and radically different time horizons in addition to their complexly conflicting interests in the present. Furthermore, they are all involved in a system that rewards those few who, like motorists who continue driving in a lane they know will be closed in a mile or so, free ride on the much larger group that obeys the rules to the letter.

It is this complex network of conflicting interests, antithetical publics, and high stakes that makes accounting what it is as a profession and as a part of the complex institutions of modern capitalism. Yet all of this is mostly absent from the book, or, perhaps better put, is simply taken for granted. That is, indeed, what most clearly marks the book as a book for accountants rather than sociologists.

I should close with some things I really liked. I suspect that many of them are old news to accountants, but they are what struck me as an outsider. I very much liked the insight that transparency can be a way of passing the buck. I like the Sennettian insight that accounting needs a public way to talk about ethical issues. I liked the discussion of "material importance," with the correlative concepts of user responsibility and *caveat lector*. I grew to like the respondents, whom I eventually started to recognize as old friends, perhaps as Gill intended by stretching their evidence out across the whole of his empirical analysis.

Gill is to be congratulated on a book that opens many of the important issues raised by accounting as a practice constitutive of the realities of modern society. But we are still waiting for a sociologically thorough theoretical and empirical analysis of those issues. This book is the starter or perhaps the fish course. The main course—perhaps it is Gill's next book—is yet to come.

REFERENCES

Jackall, R. 1988. *Moral Mazes: The World of Corporate Managers*. New York, NY: Oxford University Press.

ANDREW ABBOTT
Professor of Sociology
University of Chicago

EDITORIAL POLICY AND STYLE INFORMATION

EDITORIAL POLICY

According to the policies set by the Publications Committee (which were endorsed by the Executive Committee and were published in the *Accounting Education News*, June 1987), *The Accounting Review* "should be viewed as the premier journal for publishing articles reporting the results of accounting research and explaining and illustrating related research methodology. The scope of acceptable articles should embrace any research methodology and any accounting-related subject, as long as the articles meet the standards established for publication in the journal ... No special sections should be necessary. The primary, but not exclusive, audience should be—as it is now—academicians, graduate students, and others interested in accounting research."

The primary criterion for publication in *The Accounting Review* is the significance of the contribution an article makes to the literature. Topical areas of interest to the journal include accounting information systems, auditing and assurance services, financial accounting, management accounting, taxation, and all other areas of accounting, broadly defined. The journal is also open to all rigorous research methods.

The efficiency and effectiveness of the editorial review process is critically dependent upon the actions of both authors submitting papers and the reviewers. Authors accept the responsibility of preparing research papers at a level suitable for evaluation by independent reviewers. Such preparation, therefore, should include subjecting the manuscript to critique by colleagues and others and revising it accordingly prior to submission. The review process is not to be used as a means of obtaining feedback at early stages of developing the research.

Reviewers and editors are responsible for providing constructive and prompt evaluations of submitted research papers based on the significance of their contribution and on the rigor of analysis and presentation.

MANUSCRIPT PREPARATION AND STYLE

The Accounting Review's manuscript preparation guidelines follow *The Chicago Manual of Style* (15th ed.; University of Chicago Press). Another helpful guide to usage and style is *The Elements of Style*, by William Strunk, Jr., and E. B. White (Macmillan). Spelling follows *Webster's Collegiate Dictionary*.

FORMAT

1. All manuscripts should be formatted in 12-point font on 8 1/2 x 11″ paper and should be double-spaced, except for indented quotations.
2. Manuscripts should be as concise as the subject and research method permit, generally not to exceed 7,000 words.
3. Margins should be at least one inch from top, bottom, and sides.
4. To promote anonymous review, authors should not identify themselves directly or indirectly in their papers or in experimental test instruments included with the submission. Single authors should not use the editorial "we".
5. A cover page should show the title of the paper, all authors' names, titles and affiliations, email addresses, and any acknowledgments.
6. The American Accounting Association encourages use of gender-neutral language in its publications.

Pagination: All pages, including tables, appendices and references, should be serially numbered. Major sections should be numbered in Roman numerals. Subsections should not be numbered.

Numbers: Spell out numbers from one to ten, except when used in tables and lists, and when used with mathematical, statistical, scientific, or technical units and quantities, such as distances, weights and measures. For example: *three days; 3 kilometers; 30 years*. All other numbers are expressed numerically.

Percentages and Decimal Fractions: In nontechnical copy use the word percent in the text.

Hyphens: Use a hyphen to join unit modifiers or to clarify usage. For example: *a cross-sectional equation; re-form*. See *Webster's* for correct usage.

Keywords: The abstract must be followed by at least three keywords to assist in indexing the paper and identifying qualified reviewers.

ABSTRACT/INTRODUCTION

An Abstract of about 100 words (150 maximum) should be presented on a separate page immediately preceeding the text. The Abstract should concisely inform the reader of the manuscript's topic, its methods, and its findings. The Keywords statement should appear immediately below the Abstract. The text of the paper should start with a section labeled "I. Introduction," which provides more details about the paper's purpose, motivation, methodology, and findings. Both the Abstract and the Introduction should be relatively nontechnical, yet clear enough for an informed reader to understand the manuscript's contribution. The manuscript's title, but neither the author's name nor other identification designations, should appear on the Abstract page.

TABLES AND FIGURES

The author should note the following general requirements:

1. Each table and figure (graphic) should appear on a separate page and should be placed at the end of the text. Each should bear an Arabic number and a complete title indicating the exact contents of the table or figure. Tables and figures should define each variable. The titles and definitions should be sufficiently detailed to enable the reader to interpret the tables and figures without reference to the text.
2. A reference to each graphic should be made in the text.
3. The author should indicate where each graphic should be inserted in the text.
4. Graphics should be reasonably interpreted without reference to the text.
5. Source lines and notes should be included as necessary.

Equations: Equations should be numbered in parentheses flush with the right-hand margin.

DOCUMENTATION

Citations: Within-text citations are made using an author-year format. Cited works must correspond to the list of works listed in the "References" section. Authors should make an effort to include the relevant page numbers in the within-text citations.

1. In the text, works are cited as follows: author's last name and year, without comma, in parentheses. For example: one author, (Berry 2003); two authors, (Fehr and Schmidt 2003); three or more authors, (Dechow et al. 1998); more than one work cited, (Cole and Yakushiji 1984; Dechow et al. 1995; Levitt 1998); with two works by the same author(s), (Nelson 2003, 2005).
2. Unless confusion would result, do not use "p." or "pp." before page numbers. For example, (Dechow and Dichev 2002, 41–42).
3. When the reference list contains two or more works by the same author (as the only author or first of two or more authors) published in the same year, the suffix a, b, etc., is appended to the date in the within-text citations and in the "References" section. For example, (Johansson 2004a, 2004b, 2004c; Baiman and Rajan 2002a, 2002b; Dhaliwal et al. 2005a, 2005b).
4. When the author's name is mentioned in the text, it need not be repeated in the citation. For example: "Cohen et al. (2005) provide ..."
5. Citations to institutional works should use acronyms or short titles where practicable. For example: (NCFFR, The Treadway Commission 1987).
6. If the paper refers to statutes, legal treatises, or court cases, citations acceptable in law reviews should be used.

Reference List: Every manuscript must include a "References" section that contains only those works cited within the text. Each entry should contain all information necessary for unambiguous identification of the published work. Use the following formats (which follow *The Chicago Manual of Style*):

1. Arrange citations in *alphabetical order* according to the surname of the first author or the name of the institution or body responsible for the published work.
2. Use authors' initials instead of proper names.
3. For two or more authors, separate authors with a comma, including a comma before "and" (Dechow, P. M., R. Sloan, and A. Sweeney).
4. Date of publication follows the author's (authors') name(s).
5. Titles of journals or newspapers should not be abbreviated.

Sample entries are as follows:

Baiman, S., and M. Rajan. 2002a. The role of information and opportunism in the choice of buyer-supplier relationships. *Journal of Accounting Research* 40 (2): 247–278.
———, and ———. 2002b. Incentive issues in inter-firm relationships. *Accounting, Organizations and Society* 27 (3): 213–238.
Berry, R. 2003. Testimony before the Senate Committee on Homeland Security and Governmental Affairs Permanent Subcommittee on Investigations. November 18. Available at: http://hsgac.senate.gov/files/111803berry.pdf.
Cohen, D., A. Dey, and T. Lys. 2005. The Sarbanes Oxley Act of 2002: Implications for compensation structure and risk-taking incentives of CEOs. Working paper, New York University, University of Chicago, and Northwestern University.
Cole, R., and T. Yakushiji, eds. 1984. *The American and Japanese Auto Industries in Transition.* Ann Arbor, MI: University of Michigan.
Dechow, P. M., R. Sloan, and A. Sweeney. 1995. Detecting earnings management. *The Accounting Review* 70 (2): 193–225.
———, S. P. Kothari, and R. L. Watts. 1998. The relation between earnings and cash flows. *Journal of Accounting and Economics* 25: 133–168.
———, and I. Dichev. 2002. The quality of accruals and earnings: The role of accrual estimation errors, *The Accounting Review* 77 (Supplement): 35–59.
Dhaliwal, D., M. Erickson, and O. Li. 2005a. Shareholder income taxes and the relation between earnings and returns. *Contemporary Accounting Research* 22: 587–616.
———, L. Krull, O. Li, and W. Moser. 2005b. Dividend taxes and implied cost of equity capital. *Journal of Accounting Research* 43: 675–708.
Easton, P. 2003. Discussion of: The predictive value of expenses excluded from pro forma earnings. *Review of Accounting Studies* 8: 175–183.
Engel, E., R. Hayes, and X. Wang. 2007. The Sarbanes-Oxley Act and firms' going-private decisions. *Journal of Accounting and Economics* (forthcoming).
Financial Accounting Standards Board (FASB). 2006. *Accounting for Uncertainty in Income Taxes, and Interpretation of FASB Statement No. 109.* FASB Interpretation No. 48. Financial Accounting Series. Norwalk, CT: FASB.
Fehr, E., and K. Schmidt. 2003. A theory of fairness, competition, and cooperation. In *Advances in Behavioral Economics*, edited by C. Camerer, G. Loewenstein, and M. Rabin, 271–296. New York, NY: Princeton University Press.
Janis, I. L. 1982. *Groupthink: Psychological Studies of Policy Decisions and Fiascoes.* Boston, MA: Houghton Mifflin.
Levitt, A. 1998. The numbers game. Speech delivered at New York University, Center for Law and Business, September 28.
Maggi, G. 1999. The value of commitment with imperfect observability and private information. *RAND Journal of Economics* (Winter) 30: 555–574.
National Commission on Fraudulent Reporting (the Treadway Commission). 1987. *Report of the National Commission on Fraudulent Financial Reporting*. Washington, D.C.: NCFFR.
Nelson, M. W. 2003. Behavioral evidence on the effects of principles- and rules-based standards. *Accounting Horizons* 17 (1): 91–104.
———. 2005. A review of experimental and archival conflicts-of-interest research in auditing. In *Conflicts of Interest: Challenges and Solutions in Business, Law, Medicine, and Public Policy*, edited by D. A. Moore, D. M. Cain, G. Loewenstein, and M. H. Bazerman. Cambridge, U.K.: Cambridge University Press.
Rigdon, E. E., R. E. Schumacker, and W. Wothke. 1998. A comparative review of interaction and nonlinear modeling. In *Interaction and Nonlinear Effects in Structural Equation Modeling*, edited by R. E. Schumacker, and G. A. Marcoulides, 1–16. Mahwah, NJ: Erlbaum Associates.
Schultz, E., and T. Francis. 2002. Companies profit on workers' deaths through "dead peasants" insurance. *Wall Street Journal* (April 19): 1.
Securities and Exchange Commission (SEC). 2002. *Certification of Disclosure in Companies' Quarterly and Annual Reports.* Release Nos. 33-8124, 34-46427. Washington, D.C.: SEC.
U.S. House of Representatives. 2002. The Sarbanes-Oxley Act of 2002. Public Law 107-204 [H. R. 3763]. Washington, D.C.: Government Printing Office.

Footnotes: Footnotes are not used for documentation. Textual footnotes should be used only for extensions and useful excursions of information that if included in the body of the text might disrupt its continuity. Footnotes should be inserted using the "footnote" or "endnote" feature of the word processing software which will automatically number the footnotes throughout the manuscript with superscript Arabic numerals.

SUBMISSION OF MANUSCRIPTS

Authors should note the following guidelines for submitting manuscripts:

1. Manuscripts currently under consideration by another journal or publisher should not be submitted. The author must state upon submission that the work is not submitted or published elsewhere.
2. For manuscripts reporting on field surveys or experiments: If the additional documentation (e.g., questionnaire, case, interview schedule) is sent as a separate file, then all information that might identify the author(s) must be deleted from the instrument.
3. Manuscripts should be submitted via email as Microsoft Word or PDF file to the Senior Editor, Steven Kachelmeier, at email address tar@mccombs.utexas.edu. Please submit separate files for (1) the manuscript's title page with identifying information (not forwarded to reviewers), (2) the manuscript with title page and all other identifying information removed, and (3) any necessary supplemental files, such as experimental instructions and/or response memoranda on invited revisions. The body or subject line of the email should also indicate the submission fee receipt code from AAA (see step 4 below).
4. A nonrefundable submission fee is required unless explicitly waived by the editor for invited revisions of previous submissions. The submission fee in U.S. funds is $200.00 for members and $400.00 for nonmembers of the AAA payable by credit card (VISA or MasterCard only). The payment form is available online at: https://aaahq.org/AAAforms/journals/tarsubmit.cfm. If you are unable to pay by credit card or have any questions, please contact the AAA Member Services Team at (941) 921-7747 or info@aaahq.org.
5. Revisions must be submitted within 12 months from the decision letter inviting a revision.

COMMENTS

The Accounting Review welcomes submissions of comments on previous *TAR* articles. Comments on articles previously published in *The Accounting Review* will generally be reviewed by two reviewers, usually including an author of the original article (to assist the editor in evaluating whether the submitted comment represents the prior article accurately) and an additional reviewer who is independent of the original article. If a comment is accepted for publication, the original author will generally be invited to reply. All other editorial requirements, as enumerated above, apply to proposed comments.

POLICY ON REPRODUCTION

An objective of *The Accounting Review* is to promote the wide dissemination of the results of systematic scholarly inquiries into the broad field of accounting.

Permission is hereby granted to reproduce any of the contents of the *Review* for use in courses of instruction, as long as the source and American Accounting Association copyright are indicated in any such reproductions.

Written application must be made to the American Accounting Association, 5717 Bessie Drive, Sarasota, FL 34233-2399, for permission to reproduce any of the contents of the *Review* for use in other than courses of instruction—e.g., inclusion in books of readings or in any other publications intended for general distribution. In consideration for the grant of permission by the *Review* in such instances, the applicant must notify the author(s) in writing of the intended use to be made of each reproduction.

Except where otherwise noted in articles, the copyright interest has been transferred to the American Accounting Association. Where the author(s) has (have) not transferred the copyright to the Association, applicants must seek permission to reproduce (for all purposes) directly from the author(s).

POLICY ON DATA AVAILABILITY

The AAA's Executive Committee policy (originally adopted in 1989, and amended in 2009) is that the objective of the Association-wide journals (*The Accounting Review, Accounting Horizons, Issues in Accounting Education*) is to provide the widest possible dissemination of knowledge based on systematic scholarly inquiries into accounting as a field of professional research, and educational activity. To fulfill this objective, authors are encouraged to make their data available for use by others in extending or replicating results reported in their articles.

THE ACCOUNTING REVIEW

VOLUME 86, No. 2 MARCH 2011

A JOURNAL OF THE
AMERICAN ACCOUNTING ASSOCIATION

AMERICAN ACCOUNTING ASSOCIATION

The American Accounting Association was founded in 1916 as the American Association of University Instructors in Accounting. Its present name was adopted in 1936. The Association is a voluntary organization of persons interested in accounting education, research, and practice.

Membership in the Association is available at the following annual rates:

Full membership:
The membership dues include *Accounting Education News*, and online selection of one, two, or all three of the following Association-wide journals: *The Accounting Review*, *Accounting Horizons*, and *Issues in Accounting Education*. (This does not include Section journals.) Select from:

The Accounting Review only	U.S. $185.00	with Print $220.00
Accounting Horizons OR *Issues in Accounting Education* only	U.S. $185.00	with Print $200.00
The Accounting Review AND *Accounting Horizons* OR *Issues in Accounting Education*	U.S. $195.00	with Print $245.00
Accounting Horizons AND *Issues in Accounting Education*	U.S. $195.00	with Print $225.00
All three Association-wide journals	U.S. $205.00	with Print $270.00

Associate membership:
Full-time students are eligible for associate membership. The print option is not available for Associate members. Select from:

Any one Association-wide journal	U.S. $25.00
Any two Association-wide journals	U.S. $35.00
All three Association-wide journals	U.S. $45.00

In addition, full members may belong to one or more Sections: Accounting, Behavior and Organizations; Accounting Programs Leadership Group; American Taxation Association; Auditing; Diversity; Financial Accounting and Reporting; Forensic and Investigative Accounting; Gender Issues and Worklife Balance; Government and Nonprofit; Information Systems; International Accounting; Management Accounting; Strategic and Emerging Technologies; Teaching, Learning and Curriculum; and Two-Year College. Each Section assesses its own annual dues, which are collected by the Association.

Editorial correspondence and all manuscripts should be emailed to tar@mccombs.utexas.edu. If it is necessary to send any hard copy correspondence, the mailing address is Mary Capps, Editorial Assistant, *The Accounting Review*, The University of Texas at Austin; Department of Accounting, CBA 4M202; 1 University Station B6400; Austin, TX 78712-0211 • Telephone: (512) 471-5318 • Fax: (512) 471-3904.

The submission fee of $200.00 for members of the AAA or $400.00 for nonmembers should be paid online (VISA or MasterCard only) at: https://aaahq.org/AAAforms/journals/tarsubmit.cfm and must be paid when article is submitted. Request for reprint permission, subscriptions, advertisements, and dues should be sent to American Accounting Association, 5717 Bessie Drive, Sarasota, FL 34233-2399 • Telephone: (941) 921-7747 • Fax: (941) 923-4093 • Email: Info@aaahq.org. Notice of change of address should be mailed at least eight weeks in advance to ensure undelayed receipt of *The Accounting Review*.

The Accounting Review is indexed in the Social Science Citation Index which is included in Web of Science.

The Accounting Review (ISSN 0001-4826 print and ISSN 1558-7967 online) is published six times a year in January, March, May, July, September, and November by the American Accounting Association, 5717 Bessie Drive, Sarasota, Florida 34233-2399. General subscriptions are available at the annual rates of $395 for a print copy or at $445 for a print and electronic copy. The member subscription rate for a hard copy is $35. Electronic access is included with membership dues. Periodical postage paid at Sarasota, Florida and additional mailing offices. **POSTMASTER:** Send address change to *The Accounting Review*, 5717 Bessie Drive, Sarasota, Florida 34233-2399.

THE ACCOUNTING REVIEW

A JOURNAL OF THE

AMERICAN ACCOUNTING ASSOCIATION

VOL. 86 MARCH 2011 No. 2

BOOK REVIEWS, Stephen A. Zeff, Editor

PAST EDITORS

1926–1929	William A. Paton
1929–1943	Eric L. Kohler
1944–1947	A. C. Littleton
1948–1949	Robert L. Dixon
1950–1959	Frank P. Smith
1959–1962	Robert K. Mautz
1962–1965	Lawrence L. Vance
1965–1967	Wendell Trumbull
1968–1970	Charles H. Griffin
1971–1972	Eldon S. Hendriksen
1973–1975	Thomas F. Keller
1976–1978	Don T. DeCoster
1978–1983	Stephen A. Zeff
1983–1987	Gary L. Sundem
1987–1990	William Kinney
1990–1994	A. Rashad Abdel-khalik
1994–1997	Robert P. Magee
1997–2000	Gerald L. Salamon
2000–2002	Linda Smith Bamber
2002–2005	Terry Shevlin
2005–2008	Dan S. Dhaliwal

THE ACCOUNTING REVIEW
Vol. 86, No. 2
2011
pp. 367–383

American Accounting Association
DOI: 10.2308/accr.00000031

Accounting Scholarship that Advances Professional Knowledge and Practice

Robert S. Kaplan
Harvard University

ABSTRACT: Recent accounting scholarship has used statistical analyses on asset prices, financial reports and disclosures, laboratory experiments, and surveys of practice. The research has studied the interface among accounting information, capital markets, standard-setters, and financial analysts and how managers make accounting choices. But as accounting scholars have focused on understanding how markets and users process accounting data, they have distanced themselves from the accounting process itself. Accounting scholarship has failed to address important measurement and valuation issues that have arisen in the past 40 years of practice. This gap is illustrated with missed opportunities in risk measurement and management and the estimation of the fair value of complex financial securities. This commentary encourages accounting scholars to devote more resources to obtaining a fundamental understanding of contemporary and future practice and how analytic tools and contemporary advances in accounting and related disciplines can be deployed to improve the professional practice of accounting.

Keywords: *accounting research; professional practice; accounting education; field studies; risk measurement; pension risk; fair value measurement; options pricing model.*

I. INTRODUCTION

My talk is a response to a request from a 28-year-old, newly minted doctoral graduate soliciting advice about how to select research topics and teaching assignments for her forthcoming career as an accounting academic.

You are embarking on an exciting five-decade faculty career as a business school professor. You have acquired knowledge of contemporary accounting research issues and excellent disciplinary training in mathematics, economics, finance, statistics, and the behavioral sciences. But you should understand that you are joining the faculty of a professional school, not a graduate school or academic department. Your profession could be viewed as "accounting" or somewhat more

---

I am indebted to two HBS colleagues: Robert Merton, who inspired me with the opportunities to advance accounting scholarship and practice by building upon his pioneering work in financial economics; and Rakesh Khurana, for discussions about the role for scholarship in professional schools. I also benefited from discussions with Scott Richard, Wharton School, about the feasibility and desirability of fair value measurements for financial instruments.

Editor's note: This commentary, based on a lecture at the 2010 American Accounting Association Annual Meeting in San Francisco, CA, was invited by Senior Editor Steven Kachelmeier, consistent with the AAA Executive Committee's goal to promote broad dissemination of the AAA Presidential Scholar Lecture.

Submitted: September 2010
Accepted: October 2010
Published Online: March 2011

broadly as "management" (Khurana 2007). Among all the management disciplines, accounting is the closest to a professional discipline—many of the students we educate get certified and must maintain professional standing through continuing education programs. These accounting and financial professionals are our clients.

As a scholar in a professional school, you will educate current and future professionals on the field's common body of knowledge. You will learn how to dispense the existing common body of knowledge in ways that your students can put into action during their professional careers. All accounting and management professors have this responsibility. But also important, though less widely known, accepted, and advocated, academic scholars at a professional school should contribute to *advancing* the profession's body of knowledge, especially when innovation is high and major changes are occurring in the practice environment of the profession.

For building and sustaining a successful scholarly career in a professional school, you should be continually thinking about three fundamental questions:

1. What are the big issues faced by our practice community?
2. What are the comparative advantages that accounting scholars can bring to address these fundamental accounting and management issues? My partial list for areas in which accounting scholars should have a comparative advantage, relative to scholars in other managerial disciplines, includes measurement, reporting and disclosure, auditing, financial analysis, and management control.
3. How does our research advance knowledge in these core areas of our discipline?

II. ACCOUNTING RESEARCH: 1968–2010

During the past 40+ years, research by management and accounting scholars has produced many important insights and advanced our understanding of the environment in which accounting functions. Much of this scholarship has focused on how information and markets interact, including the information content of accounting numbers, the role of accounting accruals versus cash flows, voluntary and involuntary managerial disclosures, efficiency and anomalies in markets, and the impact of accounting choices made by managers. Some scholarship has given us insights in how financial analysts do their work and their influence on markets. I have been particularly impressed by the development of the accounting-based valuation model (Ohlson 1995; Feltham and Ohlson 1995). This is a particularly good, although somewhat rare, example of where academic accounting scholarship has created an important innovation that advanced professional practice. Academics have contributed to standard-setting practice by testing the reactions to proposed and implemented standards and communicating the implications of capital-market-based research to standard-setters. Accounting scholarship has also informed and advanced the professional work of auditors and tax professionals (AAA Research Impact Task Force 2009).

Most of these research contributions have come from statistical studies of phenomena, using archival data from financial statements, proxy statements, market prices of equity, debt, and occasionally other financial instruments, credit ratings, loan documents, bankruptcies, conference calls, and analyst forecasts. Accounting scholars have also performed statistical analysis on data collected from carefully designed laboratory experiments and simulations, and from surveys of practitioners. One characteristic of all this research is the existence of multiple tables within the article populated with estimated coefficients, at least a few of which have between one and three asterisks near them, attesting to their degree of statistical, although not necessarily economic, significance (McCloskey 1998). Scholars have also studied, both analytically and empirically, the properties of management contracts and incentives.

Table 1, extracted from the annual reports of *The Accounting Review* over the past two years (Kachelmeier 2009, 2010), summarizes the research methods used by accounting scholars in the

TABLE 1
Submissions and Acceptance to *The Accounting Review* (2008–2010)

Research Method	# Submitted	# Accepted	% Submitted	% Accepted
Empirical-Archival	933	93	75%	65%
Experimental	175	24	14%	17%
Analytic	96	18	8%	13%
Survey	30	6	2%	4%
Field and Case Study	15	2	1%	1%
Total	1249	143	100%	100%

Source: Kachelmeier (2009, 2010).

papers submitted and accepted during the two-year period 2008–2010 to the American Accounting Association's leading research journal, *The Accounting Review*. The table shows the predominance of statistical analysis on archival, mostly financial, data, and also the extensive use of laboratory experiments, surveys, and analytic models. For future reference, note the low market share (1 percent) of field and case studies.

Overall, I think this is a good record, in fact a very good record, of accomplishment. Compared to 45 years ago, we have had a great expansion of research performed and published by a broader academic accounting community, and some of this research has helped us understand, much better than in 1968, the role of accounting reporting and disclosure in markets and contracts, and how people process and use accounting information. But we are likely encountering diminishing returns in several of our research areas. The increment in our knowledge from the 371st paper on accruals versus cash flows, earnings management, or voluntary disclosure is undoubtedly much lower than the contribution from the first five to ten papers that introduced each topic into the academic accounting literature. Accounting scholars, like many of their academic colleagues, exhibit strong herding effects; they follow where others have already gone rather than forging a new path by studying a new issue in an innovative way.

III. THE GAP BETWEEN ACCOUNTING SCHOLARSHIP AND PRACTICE

If academics in professional school limit their research agenda to issues that can be adequately addressed by a narrow set of generally accepted research methods, then they must wait for phenomena to happen to generate sufficient archival data for them to analyze statistically. Thus, much of accounting (and management) research for the past 40 years has been reactive. It has studied, evaluated, and explained existing practice, but has not contributed to advancing that practice (the accounting-based-valuation model, and improvements in auditing and tax being notable exceptions). This situation would be adequate if little change or innovation was occurring in the profession. But over the past 50 years, some major discontinuities have occurred and the implications of these changes have yet to make their way into the common body of knowledge being studied by academic scholars and taught to future professionals. Among these changes are huge increases in the volume, velocity, volatility, and complexity of transactions, the globalization of markets leading to trading of assets in different time zones and currencies and to issues in reporting and managing across unclear organizational boundaries, and the introduction of new and complex securities and derivative contracts that shift risks among companies and market participants. Society has increased its expectations about corporate governance and oversight of execu-

tive compensation. A major management technology innovation has occurred for pricing risk (Black and Scholes 1973; Merton 1973), which has been overlooked and underappreciated by accounting scholars for the past four decades (much more on this later).

By responding slowly if at all to major new challenges and opportunities in the environment in which accounting is practiced, accounting scholars have become less familiar with emerging professional challenges and opportunities. Few have helped to craft new solutions and approaches for people practicing in the profession. For example, we failed to see in a timely fashion the reporting, valuation, and disclosure implications from the new types of mortgage lending and the massive securitizations of these loans.

Interestingly, accounting scholars are not the only academics who have become disconnected from practice. Paul Krugman (1994, 43–44) describes a similar situation in development economics, which he called the "hollowing of Africa" effect:

> The paper, "The Evolution of European Ignorance about Africa," describes how European maps of the African continent evolved from the 15th to the 19th centuries … The coastline of Africa was first explored, then plotted with growing accuracy, and by the 18th century that coastline was shown in a manner essentially indistinguishable from that of modern maps … On the other hand, the interior emptied out. The weird mythical creatures were gone, but so were the real cities and rivers. In a way, Europeans had become more ignorant about Africa than they had been before … improvement in the art of mapmaking raised the standard for what was considered valid data. Second-hand reports of the form "six days south of the end of the desert you encounter a vast river flowing from east to west" were no longer something you would use to draw your map. Only features of the landscape that had been visited by reliable informants equipped with sextants and compasses now qualified.

Today's accounting academics, after 40 years of using rigorous social science research methods, know much more about the coastline of accounting, its interface between accounting reports and capital markets, analysts, auditors, regulators, boards, and the media. But they know substantively less than academics of 40 years ago about the "interior" of leading-edge professional accounting and finance practice. We have experienced a "hollowing of professional practice" among accounting scholars that you and a new generation of accounting faculty could productively fill.

More than 30 years ago, Roethlisberger, a social scientist and management scholar who was a coauthor on the Hawthorne effect experiments, described the research preferences of academic social scientists to work on topics that can be explored using statistical and mathematical methods (Roethlisberger 1977, 390). He illustrated his points using the Knowledge Enterprise diagram in Figure 1.[1] Roethlisberger observed that the clinical approach of acquiring knowledge by describing and classifying the activities of skilled practitioners is frequently bypassed by researchers because clinical research is considered less elegant than deductive research.[2]

Donald Schön (1992, 54) also expressed concern that systematic study and description of professional practice holds a less exalted view in academia than deductive theory and hypothesis testing:

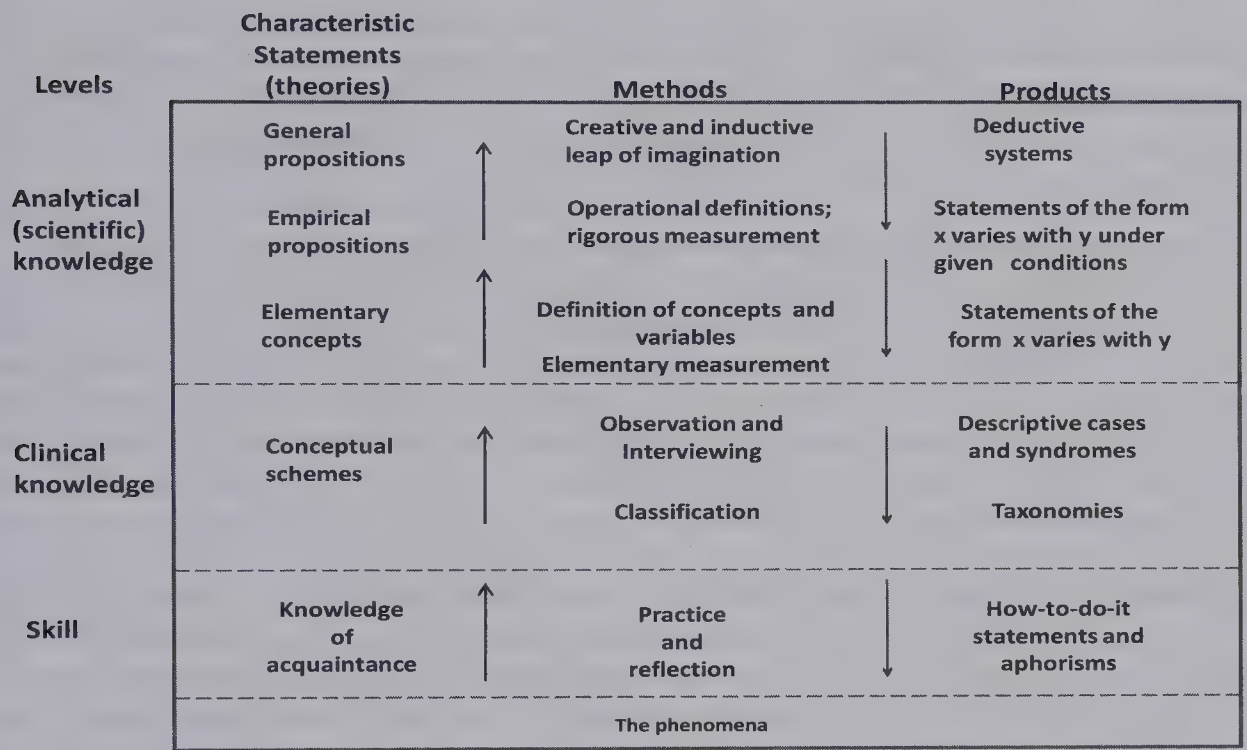
> In the varied typography of professional practice, there is a high, hard ground which overlooks a swamp. On the high ground, manageable problems [N.B. think of simple principal-agent models] lend themselves to solution through the use of research-based theory and technique. In the swampy lowlands, problems are messy and confusing and incapable of technical solution … [But] the

---

1 Summaries of the activities at the various stages of the knowledge tree—description, classification, measurement, discovering relationships, theory building, and theory testing—can be found in Kaplan (1986).

2 On academic scholars' desire for elegance, I add a comment by Einstein (1920, 5 [Preface]), "leave elegance to the tailor," which he attributed to the 19th century physicist, Ludwig Boltzmann.

FIGURE 1
Knowledge Enterprise Diagram



(For the development of knowledge, read from the bottom up; for the practice of knowledge, read from the top down.)

Source: Roethlisberger (1977, 393).

> problems of the high ground tend to be relatively unimportant to individuals or to society at large, however great their technical interest may be, while in the swamp lie the problems of greatest human concern … Some researchers … pursue an agenda driven by evolving questions of modeling theory and techniques, increasingly divergent from the contexts of actual practice.

I am less pessimistic than Schön about whether rigorous research can inform professional practice (witness the important practical significance of the Ohlson accounting-based valuation model and the Black-Merton-Scholes options pricing model), but I concur with the general point that academic scholars spend too much time at the top of Roethlisberger's knowledge tree and too little time performing systematic observation, description, and classification, which are at the foundation of knowledge creation. Henderson (1970, 67–68) echoes the benefits from a more balanced approach based on the experience of medical professionals:

> both theory and practice are necessary conditions of understanding, and the method of Hippocrates is the only method that has ever succeeded widely and generally. The first element of that method is hard, persistent, intelligent, responsible, unremitting labor in the sick room, not in the library … The second element of that method is accurate observation of things and events, selection, guided by judgment born of familiarity and experience, of the salient and the recurrent phenomena, and their classification and methodical exploitation. The third element of that method is the judicious

> construction of a theory … and the use thereof … [T]he physician must have, first, intimate, habitual, intuitive familiarity with things, secondly, systematic knowledge of things, and thirdly an effective way of thinking about things.

More recently, other observers of business school research have expressed concerns about the gap that has opened up in the past four decades between academic scholarship and professional practice. Examples include:

> Historical role of business schools and their faculty is as evaluators of, but not creators or originators of, business practice. (Pfeffer 2007, 1335)

> Our journals are replete with an examination of issues that no manager would or should ever care about, while concerns that are important to practitioners are being ignored. (Miller et al. 2009, 273)

In summary, while much has been accomplished during the past four decades through the application of rigorous social science research methods to accounting issues, much has also been overlooked. As I will illustrate later in these remarks, we have missed big opportunities to both learn from innovative practice and to apply innovations from other disciplines to important accounting issues. By focusing on these opportunities, you will have the biggest potential for a highly successful and rewarding career.

Integrating Practice and Theory: The Experience of Other Professional Schools

Other professional schools, particularly medicine, do not disconnect scholarly activity from practice. Many scholars in medical and public health schools do perform large-scale statistical studies similar to those done by accounting scholars. They estimate reduced-form statistical models on cross-sectional and longitudinal data sets to discover correlations between behavior, nutrition, and health or sickness. Consider, for example, statistical research on the effects of smoking or obesity on health, and of the correlations between automobile accidents and drivers who have consumed significant quantities of alcoholic beverages. Such large-scale statistical studies are at the heart of the discipline of epidemiology.

Some scholars in public health schools also intervene in practice by conducting large-scale field experiments on real people in their natural habitats to assess the efficacy of new health and safety practices, such as the use of designated drivers to reduce alcohol-influenced accidents. Few academic accounting scholars, in contrast, conduct field experiments on real professionals working in their actual jobs (Hunton and Gold [2010] is an exception). The large-scale statistical studies and field experiments about health and sickness are invaluable, but, unlike in accounting scholarship, they represent only one component in the research repertoire of faculty employed in professional schools of medicine and health sciences.

Many faculty in medical schools (and also in schools of engineering and science) continually innovate. They develop new treatments, new surgeries, new drugs, new instruments, and new radiological procedures. Consider, for example, the angiogenesis innovation, now commercially represented by Genentech's Avastin drug, done by Professor Judah Folkman at his laboratories in Boston Children's Hospital (West et al. 2005). Consider also the dozens of commercial innovations and new companies that flowed from the laboratories of Robert Langer at MIT (Bowen et al. 2005) and George Whiteside at Harvard University (Bowen and Gino 2006). These academic scientists were intimately aware of gaps in practice that they could address and solve by applying contemporary engineering and science. They produced innovations that delivered better solutions in actual clinical practices. Beyond contributing through innovation, medical school faculty often become practice thought-leaders in their field of expertise. If you suffer from a serious, complex illness or injury, you will likely be referred to a physician with an appointment at a leading academic medical school. How often, other than for expert testimony, do leading accounting professors get asked for advice on difficult measurement and valuation issues arising in practice?

One study (Zucker and Darby 1996) found that life-science academics who partner with industry have higher academic productivity than scientists who work only in their laboratories in medical schools and universities. Those engaged in practice innovations work on more important problems and get more rapid feedback on where their ideas work or do not work.

These examples illustrate that some of the best academic faculty in schools of medicine, engineering, and science, attempt to improve practice, enabling their professionals to be more effective and valuable to society.

Implications for Accounting Scholarship

To my letter writer, just embarking on a career as an academic accounting professor, I hope you can contribute by attempting to become the accounting equivalent of an innovative, world-class accounting surgeon, inventor, and thought-leader; someone capable of advancing professional practice, not just evaluating it. I do not want you to become a "JAE" (Just Another Epidemiologist). My vision for the potential in your 40+ year academic career at a professional school is to develop the knowledge, skills, and capabilities to be at the *leading edge* of practice. You, as an academic, can be more innovative than a consultant or a skilled practitioner. Unlike them, you can draw upon fundamental advances in your own and related disciplines and can integrate theory and generalizable conceptual frameworks with skilled practice. You can become the accounting practice leader, the "go-to" person, to whom others make referrals for answering a difficult accounting or measurement question arising in practice.

But enough preaching! My teaching is most effective when I illustrate ideas with actual cases, so let us explore several opportunities for academic scholarship that have the potential to make important and innovative contributions to professional practice.

IV. RISK MEASUREMENT AND MANAGEMENT

I (and probably many others) became aware of the gaps in risk management practice in 2007 when important institutions began to fail unexpectedly. As this trend continued through 2008, I felt a lot like Yogi Berra when he said, "It's déjà vu all over again." The 2007/08 risk management failures took me back to the early 1980s (Kaplan 1983; Johnson and Kaplan 1986), when we learned, to our surprise, that cost accounting and performance measurement, highly important accounting functions, were being done poorly by many enterprises.

Risk management is a great issue for accounting academics. It builds upon our fundamental comparative advantages as an academic discipline: expertise in measurement, valuation, reporting, disclosure, management control, auditing, and governance. The topic contains issues relevant for financial reporting, management control, and auditing, three of our most important subject areas in the American Accounting Association.

Among the risk management issues I would like to know more about, consider the following questions:

1. How can we measure or quantify risk? Measurement is about the past; even so-called leading indicators are measuring events that have already occurred. How can we quantify risk or develop risk indicators for an event that has not yet occurred and, we hope, may never occur? Quantifying risk exposure is a challenging measurement issue.
2. Will a single risk management system work for all types of risks, or do different types of risks require different types of risk management systems?
3. The optimal level of risk is not zero. How can companies introduce risk management without sacrificing their innovation and risk-taking activities? What is the appropriate balance between innovation and risk management, and how can this balance be maintained?

4. Much work is currently being done by regulators and standard-setters (COSO, Basel, rating agencies, stock exchanges, etc.) to promulgate rules and standards on companies' risk management practices:

   a. Is the practice of risk management sufficiently stable, mature, and understood that now is a good time to develop risk standards and regulations? Or is it better for companies to innovate and experiment with different risk management approaches before regulators standardize and codify practices?
   b. How can academic scholars participate in and document such innovation and experimentation?
   c. Does and should the optimal form of risk management vary by type of company, strategy, degree of innovation? How?
5. What is the relationship between a professional risk management officer and line management?
6. What is the relationship between risk management and internal audit?
7. Other than perhaps in financial institutions, what do people mean when they talk about "risk appetite"? When the Board is told it must determine the "risk appetite" for the company, how can risk appetite be quantified and made operational?
8. Can scholarship improve the reporting and disclosure of risk in company documents?

On this last question, I looked at a recent risk disclosure in a company's 10-K:

> In the event of changes in market conditions, such as interest or foreign exchange rates, equity, fixed income, commodity or *real estate valuations*, liquidity, availability of credit or volatility, our business could be adversely affected in many ways … Further declines in real estate values in the U.S. and continuing credit and liquidity concerns could further reduce our level of mortgage loan originations and increase our mortgage inventory while adversely affecting its value. (emphasis added)

Does this sound like the risk exposure of a huge financial institution that would file for bankruptcy less than two months after this 10-K submission? Yet this was the "risk disclosure" in the 2008 second quarter filing of Lehman Brothers, a financial institution born in the South in the 1850s. Lehman survived the U.S. Civil War, World War I, the Great Depression of the 1930s, and World War II. It built its capital during the great post-World War II global expansion, and somehow failed after a 5 percent decline in U.S. real estate prices. After 40+ years of academic research on capital markets and financial economics, is Lehman's 10-K disclosure the best we can offer to quantify and disclose a company's risk exposure? I hope not.

At present, we do not know enough to address the risk management questions with research methods at the top of the "knowledge enterprise." If we rely on traditional social science research methods, we will provide sketchy answers and only with long lags. Risk management in organizations is highly complex and context-specific. It involves many moving and interconnected components including leadership, organizational structure, incentives, finance, management control, and … risk and uncertainty. We are going to learn little by sitting in our offices and analyzing data that others produce. Similar to how we went about developing new approaches to the cost and performance measurement issues in the 1980s, we should start by using research methods at the base of the knowledge enterprise tree: identify, observe and describe the practice leaders, document and analyze the practice innovations, develop the associations of where risk management seems to be working well and where it has failed, and start to formulate the general principles underlying effective risk management practice.

While this research agenda will take years of work, I can describe several accounting topics in risk measurement that could be immediately addressed by creatively applying existing knowledge.

Measuring Pension Plan Risk

Pension accounting has been addressed by standard-setters for decades, yet little has been done on measuring and reporting the risk of pension plans. Measuring the risk of pension plans is important not only for companies with defined benefit pension plans, but also for states, municipalities, and even countries. Consider an example, adapted from Merton (1995, 473), of two companies, each with $100 million in net present value of pension liabilities (incidentally, the calculation of pension liabilities is a measurement that accountants have already out-sourced to actuaries):

> Company A holds $90 million in default-free government bonds whose durations are matched perfectly with the liabilities in its pension plan. Company B holds $25 million in such duration-matched government bonds plus $75 million in a diversified portfolio of common stocks. Which pension plan is riskier?

Today's accounting standards would report that A's plan is riskier because B's plan is "fully funded" while A's is underfunded by $10 million. But Company B holds the equivalent of a giant derivative. It could have held $100 million in duration-matched risk-free assets, but it made the equivalent of a $75 million total return on equity-for-debt swap. Now, we all know the equity risk premium is about 800 basis points higher return per year than the return from risk-free bonds. Compounded over many years, a diversified portfolio of common stocks should return far more than the risk-free portfolio. So how risky could B's equity-for-debt swap be?

Suppose the trustees of B's plan conduct the following thought experiment. Even though stocks are expected to out-perform bonds over long periods of time, how much would it cost to purchase insurance to guarantee that the return on the equity portfolio, 20 years from now, is not lower than what could be achieved in a zero-risk government bond portfolio? While I have not independently calculated the following, an expert in this field has estimated to me that such insurance would cost $15 million or more. So it would cost plan B's trustees at least $15 million to protect its asset holding from generating returns below the risk-free rate. In fact, Company B's plan is significantly riskier than Company A's because of its equity-for-debt swap.

Today, however, such quantification of this risk exposure appears nowhere in the company's financial statements even though the technology for assessing the risk of Company B's plan has been around for four decades, ever since the introduction of the options pricing model. How do investors and creditors get information about the self-insurance risk that companies (and states and countries) take on when their pension plans contain risky assets intended to pay the commitments from their fixed pension liability structure?

An accountant attempting to calculate the risk of a pension plan's assets would need to decide which options model, among many alternatives, to use. Merton describes that an options pricing model is like an autopilot on an airplane. Both process a great deal of information in a coherent and consistent way—perhaps better than human intuition and experience. But pilots understand the model and assumptions being used by the autopilot mechanism and are prepared to override them based on judgment and experience. Which accounting scholars have acquired the experience and judgment to know when and which options pricing models are valid and under what circumstances to override the model? A model that is highly accurate for short periods of time may have high errors for long-term valuation, while one that is approximately valid for long periods may be quite inaccurate for short time intervals. Merton posed the following challenge to the accounting profession:

> accounting does a good job at valuation but … it is totally inadequate to deal with risk allocation, which … is one of the critical issues today. Accounting must change in a major way to address this in the future. (Merton 1995, 470)

His challenge, however, has gone unaddressed. The entire focus of the FASB and IASB, and almost all accounting academics, has been to debate the valuation of assets and liabilities but not their risk.

What would a risk balance sheet look like? I do not know; if I did, I would be writing about it already. Yet, this is exactly the problem of not having scholarship follow contemporary practice. The technology to develop a risk balance sheet has existed for decades. To get started learning about this issue, I would attempt to learn from the most skilled practitioners I could find, probably in hedge funds, investment banks, and sophisticated diversified corporations whose financial officers must assess and manage every single day their risk exposure from mismatched asset and liability positions. This is a great issue where you could establish a reputation as the world's "risk balance sheet practice innovator and thought-leader." If successful, your framework would be taught throughout this century to accounting and business students, and would also be applied widely in practice.

Studying the Role for Credit Default Swaps

A credit default swap (CDS) represents another practice innovation that has profound implications for accounting and auditing practice. Introduced by J.P. Morgan in the mid-1990s, a CDS, similar to a deep out-of-the-money put option, offers an insurance policy on the default of risky debt. An increase in a CDS rate implies an increased likelihood of future asset price declines or an increase in the volatility of future cash flows. Accounting academics have conducted much accounting research on assessing credit and bankruptcy risk, yet have paid little attention to this new security that provides a pure signal on default likelihood. The CDS signal avoids the complications from bond or credit-specific features—such as sinking funds, collateral, and covenants—that make inferring company risk from bond yields a complex calculation.

Governments and legislators continue to debate the role of rating agencies, and how to improve their performance. The U.S. government just passed a 2,000-page bill to increase regulation of financial markets hoping to avoid future systemic risk from the failure of one or more large financial institutions. Many of the details for its implementation have been left to various regulatory agencies. Who do you think would do a better job of assessing the default risk of a major financial institution? A government regulator, who is paid less than a new assistant professor of accounting, or the aggregation of beliefs among sophisticated professionals betting hundreds of millions of dollars, and more, on understanding and pricing the risk of corporate default?[3] (This is, I hope, a rhetorical question.)

Accounting professionals can play an important role to improve the pricing of CDSs by their reporting and disclosure of information about all the debt, commitments, and contingencies incurred by a company (or public sector entity such as a state and country). Among the interesting questions, about which we currently lack adequate knowledge on the quality and reliability of CDS pricing, are the following:

1. Are CDS prices a valid estimate of default likelihood? Do they predict financial distress better than our current models of bankruptcy and credit risk? These questions must be answered affirmatively before we could consider shifting corporate risk measurement from bond-rating agencies and regulators to CDS securities markets.
2. What information is impounded in a CDS price? What new information causes a CDS price to change?[4]

---

[3] I am indebted to Professor Scott Richard, of the Wharton School, for this provocative suggestion.

[4] I am encouraged to see an article in *The Accounting Review* (Callen et al. 2009) begin to address this topic, but this first CDS paper in an accounting journal appeared five years after they started appearing in finance journals, and 15 years after the original innovation itself.

3. What kinds of risk models do sophisticated participants in CDS markets use?

For you to become the world's thought-leader on the pricing and information content of credit default swaps, I do not recommend using only reduced-form statistical models of the correlations between CDS prices and various accounting and financial ratios. CDS pricing is an emerging practice area, and I would attempt to learn, through observation and description, about the models and processes used by leading CDS-pricing practitioners. You could then move on to classify the various approaches you encountered and conduct association studies to assess the performance of these practice-based models against alternative default prediction models, or the human judgment of bond-rating personnel and government regulators.

V. FAIR VALUE MEASUREMENT

A third opportunity for innovative and important faculty research and scholarship is the measurement of fair values. Unlike research on pension plan risk and CDS pricing, many accounting scholars are already doing correlational studies between fair values and security prices, motivated to a considerable degree by the FASB and IASB mandates for increased fair value measurement and disclosure. This research reinforces the benefits from epidemiological scholarship on contemporary practice issues. Yet all this scholarship takes the "fair values" as given and does not explore how fair values actually get estimated. The accounting profession is in danger of potentially outsourcing this critical measurement, and its validation, to others. Do we have, today, academic experts on how to do "fair value" analysis for complex assets? Who is our leading academic "fair value" thought-leader, the go-to person when a company or audit firm has a difficult question on a complex fair value calculation with a lot at stake?

Let us explore the state of the art today on determining fair values for financial instruments, which are, by far, the simplest applications for fair value measurements. The traditional accounting approach of using contemporary market prices works well for assets that trade continually in thick markets. For assets that are not actively traded, banks advocate and accounting educators teach the discounted cash flow approach, using the interest rate at the time the financial asset was issued. While current accounting standards require that impairments in these assets get recognized, most banks argue against recognizing impairments as long as debtors continue to make payments. This leads to nontraded or thinly traded financial assets being carried at historical cost (or terminal value). Wachovia, and other banks, resisted fair value reporting of their financial assets by classifying them into their "held-to-maturity" portfolios, a classification that defies economic substance except in a highly restricted case. Wachovia in July 2008, reported \$75 billion in shareholders' equity, even after taking "modest" impairments of more than \$10 billion during the previous 12 months in its more than \$300 billion loan portfolio (valued at historical cost). Yet less than three months later, the bank had failed, and its acquirer, Wells Fargo, wrote down Wachovia's asset loan position by an additional \$74 billion.

This incident, and many others at the time, reveals a major shortcoming in the contemporary financial reporting framework. The deterministic discounted cash flow model is not adequate for estimating the fair values of risky financial assets. And, sadly, the ability to estimate fair values of thinly traded financial assets has existed for decades.

Valuing a Risky Financial Debt Instrument

Again, I am in debt to my finance colleague Robert Merton, who proposed a simple (some would call it "elegant") model, based on the following equation, to calculate the value of risky debt (Merton 1974):

Risk-free debt = Risky debt + insurance policy on the value of the underlying asset.

This equation states that a holder of risky debt can transform the asset into risk-free debt by purchasing an insurance policy that pays off should the debt default (the insurance could be a

default-free CDS, helping to explain the popularity of this financial instrument). After rearranging terms in the above equation, we obtain:

Risky debt = Risk-free debt

– insurance policy that the debtor ' s assets exceed the loan amount.

In other words, a bank lending to a consumer, commercial customer, or homeowner holds the equivalent of a risk-free debt position less a short position on an insurance policy that the debtor defaults. The following example makes this insight clearer:

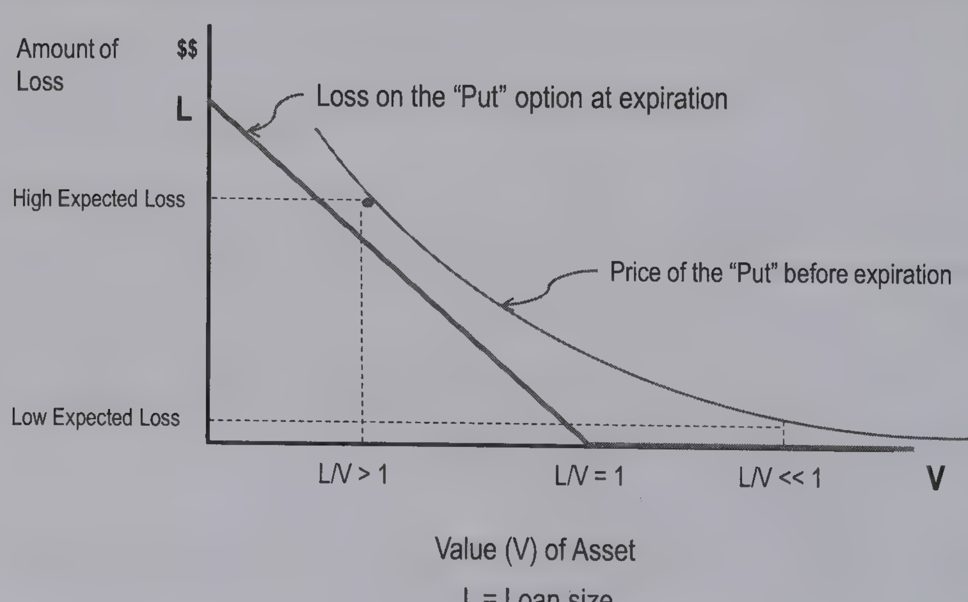
> Consider a bank that makes a loan, L, which is secured by the assets of a company or homeowner currently having a value, V. The debtor now owns a put option for the face value of the loan because at the loan's maturity:

If $V > L$, the company (homeowner) repays the loan and banker gets back L.

If $V < L$, the company (homeowner) defaults and the bank loses $L - V$. Figure 2 shows the graphical representation of the bank's loss. The downward sloping line is the loss when the terminal value is less than the loan value. The horizontal portion shows that the loss will be zero if the asset value at loan maturity exceeds the loan value.

The hyperbola above this line in Figure 2 shows the value of the bank's short position in the put option that is embedded in this loan, for a time period before the loan's maturity. When V is much greater than L, the put option is deeply out of the money and there is essentially no decline in the fair value of the loan. But as the asset value approaches the loan value from the right (when V is only slightly greater than L), the put option increases sharply in value. Any potential purchaser of the loan will be estimating the likelihood of further declines in the asset's value and also

FIGURE 2
Graphical Representation of a Bank's Potential Loss on a Secured Loan



Source: Merton (2009).

the increased volatility of the asset's price, which occurs as prices decline. In deciding how much to offer to purchase the risky loan, the potential purchaser will subtract the value of the put option from the terminal loan value (L), even though the loan may still be current (and therefore not impaired at all, as argued by the bank).

The large "bid-ask" spread, between what a rational bidder is prepared to pay for a risky loan with a loan-to-value ratio of 1, or less, is interpreted (incorrectly) by the bank as a decrease in the liquidity for its asset, not the actual decline in its underlying value. The bank rejects the markdown to fair value (the bid price) by claiming (1) it intends to hold the loan to maturity, and (2) the underlying cash flows from the asset have yet to be impaired in any way. But these are not valid reasons for failing to fair value a financial asset. When a company leverages its asset position with debt that is shorter term than the underlying asset, it loses control over whether the asset will be held to maturity; that right has been transferred to its creditors. They will make their own estimates of the asset's fair value and will likely subtract an additional discount to compensate for the possibility (from lack of disclosure) that they are dealing with a "lemons problem." Should the creditors' estimates of the bank's asset values be less than the bank's current liability position, the market for the bank's credit will disappear precipitously. The bank will complain about a sudden loss in liquidity, but the loss has been in the value of its asset holdings, not in the liquidity of markets. It is in the bank's self-interest to understand the fair value of its asset position better than its creditors, which many fail to do.

Fair Values are Often Not the Most Recent Market Price

A secured loan, illustrated above, is just one example of a thinly traded financial asset. Accountants could benefit from learning how to use contemporary market prices of similar assets to infer what price a particular asset would trade at should it trade at the next point of time. In the case of mortgage-based assets in the second half of 2008, trading may have decreased, but it was never zero. Nearly 1,000 securities traded each day, and these provided sufficient reference points to value all the asset holdings. A similar situation occurs for U.S.-based mutual funds holding equity assets that trade exclusively in overseas markets. The funds must, at the end of the U.S. trading day, calculate the fair values of these overseas equities for calculating the price at which it will buy or sell fund shares. The funds use estimates of the correlations between overseas and U.S. stock prices as well as the information in the closing prices of U.S.-traded equities to estimate the price at which the overseas equities *would have traded* in a transaction executed at the U.S. closing time.

As another example, U.S.-based funds of fixed income assets hold individual securities, most of which do not trade continually or even daily. Yet, at the end of each trading day, they must mark all their holdings to best estimates of their fair values. This calculation requires extensive statistical analysis of the correlations between the assets held by the funds and the assets that did trade at or near the end of the day. In effect, the funds use models similar to those used by the website zillow.com to estimate home prices based on actual sales of homes in the same neighborhood. The website has an algorithm that adjusts the actual transaction prices of homes in the neighborhood for differences between the traded homes and the nontraded homes. Financial institutions, such as Bear-Stearns, AIG, Merrill Lynch, and Lehman Brothers, that had not developed the capabilities to estimate the fair value of their asset holdings at the end of each trading day ended up being surprised by their insolvencies in 2007/08.

Fair Value of Compensation Packages

Complex fair value calculations also arise in executive compensation contracts, as illustrated by the "inducement portion" of the compensation plan offered to Carol Bartz to become CEO of Yahoo:

Shares Granted	Vesting Price	% Increase over Current Market Price
5,000,000	11.73	0%
1,666,667	17.60	50%
833,333	20.53	75%
833,333	23.46	100%
416,667	26.39	125%
416,667	29.33	150%
833,333	35.19	200%

Source: Yahoo Form 10-K for period ending 12/31/08. Note 17: 109.

Accounting scholars have studied compensation plans for decades. Has this research helped us to understand the properties of Bartz's inducement grant? Could we advise a member of the Yahoo! Compensation committee about the parameters of such a grant, such as the number and distribution of price points and the quantity of shares vested at each price point? How many accounting scholars or our students could calculate the fair value of this equity grant? If not, how can we give guidance to our graduates who we hope will become the auditors, CFOs, and board members of companies that issue such compensation contracts? Given all our interest in CEO compensation, this seems like it could be a fruitful area for our research and teaching.

I conducted some field research while at the 2010 AAA Annual Meeting to ascertain how we explain fair value measurements today to accounting students. I walked around the exhibition hall to examine Intermediate Accounting textbooks (a task that enabled me to skip my weight machine exercises in the hotel's fitness center later that day). In skimming the fair value chapter in each book, I saw detailed instructions on how to record asset value changes to fair values as required by current standards and regulations, but no coverage of how to estimate the fair values of risky financial assets. We have, implicitly, delegated to others the task of estimating fair values of complex securities and compensation contracts. Our students learn only how to enter the debits and credits, not the knowledge or expertise to estimate fair values. I doubt whether accounting professionals who have outsourced such a critical measurement can understand or audit the validity of the fair value estimates supplied to them.

I left the exhibition hall rather depressed. I imagined what a comparable situation would have been in physics in 1945, 40 years after the development of quantum mechanics. Would academic physicists still be teaching only Newtonian mechanics, and not instructing their students that this theory was not even approximately valid in certain important situations? Would biology professors in 1995 be instructing students only about 19th century Linnaean classification and Darwinian evolution, and ignoring a half-century of innovations in genetics and molecular biology?

I learned earlier this year that a major bank employs 500 accountants to mark its entire global portfolio of securities to fair value each day. The chief accounting officer told me that the bank cannot hire graduates from U.S. accounting departments for this task. The students do not know sufficient economics, mathematics, and statistics to perform the fair value calculations. This deficit is a direct result of accounting scholars not doing research on fair value measurement and therefore not being able to teach our students how to perform such calculations. And, unfortunately, measuring the fair value and risk of financial instruments is easy compared to estimating the fair value and risk of a company's real and intangible assets, including the capabilities and loyalty of its employees, its network of global suppliers, its customer relationships, and its new product pipeline.

VI. IMPLICATIONS FOR ACCOUNTING SCHOLARSHIP

To my young accounting scholar, I hope that you can begin to rectify this situation by attempting to increase the relevance and impact of your research, education, and teaching. Indi-

vidual faculty scholarship, however, is only one component for closing the gap between academics in professional schools and their audience of current and future practitioners. Academic scholarship is a system of interconnected parts. Reform requires changes in all the system's elements including:

1. Doctoral training (Polzer et al. 2009).
2. Publication standards applied by journal editors and reviewers.
3. Faculty promotion criteria.
4. Textbook writing and educational curriculum.
5. School and program accreditation criteria.
6. University promotion standards that are sensitive to the different expectations between scholars in academic versus professional schools.

Over the past 40 years, these six elements have converged into a stable equilibrium of accounting scholarship characterized by a narrow set of acceptable research methods, all of which are clustered at the top of Roethlisberger's (1977) Knowledge Enterprise diagram. It will require great energy to break out of this equilibrium to allow and reward research methods that document, classify, and create new knowledge for emerging problems and opportunities in practice. That is not your job as a new Assistant Professor; it is a task that falls on tenured Full Professors in your profession to take on.

Innovation in Accounting Education

In my remaining time, let me address just component #4 in the system of academic scholarship: the implications of my remarks for your teaching and educational activities. First, take advantage of being in a professional school by choosing a portion of your teaching assignment to be in executive education programs. As you develop expertise in a particular field relevant to the professional practice of accounting and finance, you can test your ideas by teaching them to experienced professionals. These executives are sacrificing much of their time and their company's money to attend and benefit from your teaching. By sharing your research with them, you get excellent feedback along two dimensions:

1. Is the problem you are addressing relevant to their experience and practice?
2. Is the expertise you have developed for this problem relevant to their practice? Are you teaching them something worthwhile about the problem that they can productively apply to their company's situation?

If the answer to either of these two questions is "no," then you will certainly hear about their disappointment. But if knowledgeable professionals value learning about your proposed solution to a problem they face, you will have accomplished an important and highly satisfying teaching milestone. And you will not have to rely only on large sample surveys to get feedback on the relevance, applicability, and innovation of your proposed solution to an important problem in practice.

Second, mastering some of the new common body of knowledge for future accounting and finance professionals will require your students to receive extensive training in economics, finance, higher mathematics (including stochastic calculus), and multivariate statistics. This may be beyond what can be delivered in a core curriculum in undergraduate accounting or a two-year M.B.A. program. But certainly, as an innovative accounting scholar, you can develop teaching materials, and new courses for an elective track in valuing and measuring the risk for complex assets, strategies, and transactions. You will contribute to your profession by educating future practitioners in innovative, contemporary solutions to problems they will encounter during their careers.

Third, try to codify your innovations by developing new classroom teaching materials and innovative textbooks (or at least textbook chapters) that will allow your insights to be taught by

your accounting academic colleagues to a much wider audience than your own students. This will enable your contributions to have a wider impact on future accounting practice.

VII. CONCLUSION

I hope you can now appreciate the many opportunities for productive scholarship during the next four or more decades of your academic career. Risk measurement and management is only one example of the many interesting and important issues being faced by accounting and finance professionals today. I am sure that the next four decades will contain as many new developments and challenges as the previous four have uncovered. I regret only that I cannot turn my own biological clock back 40+ years so I could join with you on the wonderful journey ahead.

While on this journey, work on important problems whose solutions will expand the common body of knowledge for the practice of accounting, finance, and management. Use research methods that help you understand the problems professionals face and attempt to develop innovative solutions that they can apply. Teach your insights and solutions to your students, and share them widely with your global academic colleagues. Remain excited about the research and teaching opportunities in your chosen academic field. Help to improve the world of professional practice.

REFERENCES

AAA Research Impact Task Force. 2009. The impact of academic accounting research on professional practice: An analysis by the AAA Research Impact Task Force. *Accounting Horizons* 23 (4): 411–456.

Black, F., and M. Scholes. 1973. The pricing of options and corporate liabilities. *The Journal of Political Economy* 81 (3): 637–654.

Bowen, K., A. Kazakis, A. Muir-Harmony, and B. Lapierre. 2005. The Langer Lab: Commercializing science. Harvard Business School Case #605-017. Cambridge, MA: Harvard University Press.

Bowen, K., and R. Gino. 2006. The Whitesides Lab. Harvard Business School Case #606-064. Cambridge, MA: Harvard University Press.

Callen, J. L., J. Livnat, and D. Segal. 2009. The impact of earnings on the pricing of credit default swaps. *The Accounting Review* 84 (5): 1363–1394.

Einstein, A. 1920. *Relativity: The Special and General Theory*. London, U.K.: Methuen & Co Ltd.

Feltham, G., and J. A. Ohlson. 1995. Valuation and clean surplus accounting for operating and financial activities. *Contemporary Accounting Research* 11 (2): 689–731.

Henderson, L. J. 1970. *On the Social System*. Chicago, IL: University of Chicago Press.

Hunton, J. E., and A. Gold. 2010. A field experiment comparing the outcomes of three fraud brainstorming procedures: Nominal group, round robin, and open discussion. *The Accounting Review* 85 (May): 911–935.

Johnson, H. T., and R. S. Kaplan. 1986. *Relevance Lost: The Rise and Fall of Management Accounting*. Boston, MA: Harvard Business School Press.

Kachelmeier, S. J. 2009. Annual report and editorial commentary for *The Accounting Review*. *The Accounting Review* 84 (November): 2047–2075.

———. 2010. Annual report and editorial commentary for *The Accounting Review*. *The Accounting Review* 85 (6): 2173–2203.

Kaplan, R. S. 1986. The role for empirical research in management accounting. *Accounting, Organizations and Society* 11 (4–5): 429–452.

———. 1983. Measuring manufacturing performance: A new challenge for managerial accounting research. *The Accounting Review* 58 (4): 686–705.

Khurana, R. 2007. *From Higher Aims to Hired Hands: The Social Transformation of American Business Schools and the Unfulfilled Promise of Management as a Profession*. Princeton, NJ: Princeton University Press.

Krugman, P. 1994. The fall and rise of development economics. In *Rethinking the Development Experience*, edited by L. Rodwin, and D. Schön, 39–58. Washington, D.C.: The Brookings Institution.

McCloskey, D. N. 1998. The rhetoric of significance tests. In *The Rhetoric of Economics*, 2nd edition, 112–138. Madison, WI: University of Wisconsin Press.

Merton, R. C. 1973. Theory of rational option pricing. *The Bell Journal of Economics and Management Science* 4 (1): 141–183.

———. 1974. On the pricing of corporate debt: the risk structure of interest rates. *The Journal of Finance* 29 (2): 449–470.

———. 1995. Financial innovation and the management and regulation of financial institutions. *Journal of Banking & Finance* 19 (3–4): 461–481.

———. 2009. Observations on macro financial risk propagation, structural risks, and regulatory recommendations from the crisis. Harvard Business School Deep Dive Seminar Series. (January 23). Cambridge, MA: Harvard University Press.

Miller, D., R. Greenwood, and R. Prakash. 2009. What happened to organization theory? *Journal of Management Inquiry* 18 (4): 273–279.

Ohlson, J. A. 1995. Earnings, book values, and dividends in equity valuation. *Contemporary Accounting Research* 11 (2): 661–687.

Pfeffer, J. 2007. A modest proposal: How we might change the process and product of managerial research. *Academy of Management Journal* 50 (6): 1334–1345.

Polzer, J. T., R. Gulati, R. Khurana, and M. Tushman. 2009. Crossing boundaries to increase relevance in organizational research. *Journal of Management Inquiry* 18 (4): 280–286.

Roethlisberger, F. J. 1977. *The Elusive Phenomena*. Cambridge, MA: Harvard University Press.

Schön, D. A. 1992. The crisis of professional knowledge and the pursuit of an epistemology of practice. *Journal of Interprofessional Care* 6 (1): 49–63.

West, J., A. Sagar, and M. Ashiya. 2005. Judah Folkman and the war on cancer. Harvard Business School Case #604-091. Cambridge, MA: Harvard University Press.

Zucker, L., and M. Darby. 1996. Star scientists and institutional transformation: Patterns of invention and innovation in the formation of the biotechnology industry. *Proceedings of National Academy of Sciences USA* 93 (23): 12709–12716.

THE ACCOUNTING REVIEW
Vol. 86, No. 2
2011
pp. 385–416

American Accounting Association
DOI: 10.2308/accr.00000027

Investor Trading and the Post-Earnings-Announcement Drift

Benjamin C. Ayers
The University of Georgia

Oliver Zhen Li
The University of Arizona

P. Eric Yeung
The University of Georgia

ABSTRACT: We examine whether the two distinct post-earnings-announcement drifts associated with seasonal random-walk-based and analyst-based earnings surprises are attributable to the trading activities of distinct sets of investors. We predict and find that small (large) traders continue to trade in the direction of seasonal random-walk-based (analyst-based) earnings surprises after earnings announcements. We also find that when small (large) traders react more thoroughly to seasonal random-walk-(analyst-) based earnings surprises at the earnings announcements, the respective drift attenuates. Further evidence suggests that delayed small trades associated with random-walk-based surprises are consistent with small traders' failure to understand time-series properties of earnings, whereas delayed large trades associated with analyst-based surprises are more consistent with a longer price discovery process. We also find that the analyst-based drift has declined in recent years.

Keywords: *post-earnings-announcement drift; earnings expectations; market efficiency; small and large traders; investor sophistication.*

Data Availability: *All data are publicly available.*

JEL Classifications: *G14; M41.*

We thank Steve Baginski, Linda Bamber, Mei Cheng, Ted Goodman, two anonymous referees, and workshop participants at the Chinese University of Hong Kong, Hong Kong University of Science and Technology, Hong Kong Polytechnic University, City University of Hong Kong, University of Hong Kong, Nanjing University, Nanyang Technological University, National University of Singapore, New York University, the Southwestern University of Finance and Economics (China), The University of Arizona, University of Connecticut, University of Oregon, and the 2009 American Accounting Association Financial Accounting and Reporting Section Midyear Meeting for their helpful comments. Professor Ayers gratefully acknowledges the financial support from the Terry College of Business. Professor Li gratefully acknowledges the financial support from the Eller College of Management. Professor Yeung gratefully acknowledges the financial support from the Terry-Sanford Research Awards and the computing support from the Research Computing Center at The University of Georgia.

Editor's note: Accepted by Beverly Walther.

Submitted: March 2009
Accepted: August 2010
Published Online: March 2011

I. INTRODUCTION

More than four decades of research on post-earnings-announcement drift consistently finds that stock prices tend to move in the direction of earnings surprises following earnings announcements. While the majority of prior studies focus on the drift associated with seasonal random-walk-based earnings surprises, recent studies document a drift associated with analyst-based earnings surprises that appears not only distinct from, but also even larger than the drift associated with seasonal random-walk-based earnings surprises (Livnat and Mendenhall 2006; Doyle et al. 2006). Despite the extensive literature on post-earnings-announcement drift, there is no consensus to date regarding the source(s) of either drift.[1]

Our study examines whether the distinct drifts are attributable, at least in part, to different identifiable subsets of investors who under-react to different forms of earnings innovations. Evidence on this issue is important for researchers and practitioners to understand the nature and the source of distinct drifts associated with seasonal random-walk-based earnings surprises and analyst-based earnings surprises. Indeed, Livnat and Mendenhall (2006, 180) argue that "[i]f researchers do not understand how the magnitude of the drift depends on the specification of earnings surprise, they stand little chance of understanding the nature of the anomaly."

We hypothesize that the two post-earnings-announcement drifts are attributable, at least in part, to distinct sets of investors (small and large traders) who use different earnings expectations models, but fail to fully react to earnings news. Specifically, if small traders use seasonal random-walk-based earnings expectations models but under-react to the seasonal random-walk-based earnings surprises, then we expect that post-announcement trading by small traders helps to explain the drift associated with seasonal random-walk-based earnings surprises (hereafter, RW drift). Similarly, if large traders base their earnings expectations on analyst forecasts but do not fully react to the analyst-based earnings surprises, then we anticipate that post-announcement trading by large traders helps to explain the drift associated with analyst-based earnings surprises (hereafter, AF drift).

Despite the intuitive appeal of our expectations, there is little evidence that links either drift to specific investors' trading. For example, although Battalio and Mendenhall (2005) and Bhattacharya (2001) find that small traders are more likely to trade on the RW earnings surprise around earnings announcements and large traders trade on the AF earnings surprise around earnings announcements, they do not link the RW or AF *drift* to trades by specific investor groups in the post-announcement period. Bartov et al. (2000) find that the RW drift decreases with the level of institutional ownership and conclude that institutional investors improve the efficiency with which RW earnings surprises are priced. However, they do not investigate the AF drift or link investor trades to the RW drift. More recently, Hirshleifer et al. (2008) investigate whether the RW drift is attributable to individual investors trading *contrary* to the RW earnings surprises at earnings announcements, which creates underpricing after good news and overpricing after bad news. Contrarian trading by individual investors is an alternative sufficient, but not *necessary* condition for the RW drift. Hirshleifer et al. (2008) find no evidence that individual investors trade contrary to RW earnings surprises, and offer three alternative explanations for the RW drift: the RW drift is an artifact of poor measurement of risk-adjusted returns; individual investors cause the RW drift but their proprietary sample from a single discount brokerage fails to detect this relation; or institutional investors cause the RW drift. In sum, the effect of distinct investor groups on the RW

[1] For example, prior research conjectures that the drift associated with seasonal random-walk-based earnings surprises may be attributable to investor under-reaction (Bernard and Thomas 1989; 1990), a subset of naïve investors trading contrary to earnings news (Hirshleifer et al. 2008), other complex short- and long-term trading patterns by investors (Bartov et al. 2000; Ke and Ramalingegowda 2005), and mismeasurement of risk-adjusted returns (Ball 1992; Kothari 2001).

drift is far from settled and there is little explanation as to whom or what causes the AF drift.

To build on prior research, we investigate the possibility that the RW and AF drifts could be explained by distinct groups of traders who under-react to differing earnings news. Using trade and quote data for trades executed from 1993 through 2005, we classify investors as small or large traders based on trade size, and use a buy-sell order imbalance metric to estimate large and small traders' net buying activities. We then examine small and large trades around and after earnings announcements and their relation to the post-earnings-announcement drifts.[2]

Results are consistent with our predictions. We find that only small traders systematically trade in the direction of the RW earnings surprises after earnings announcements. Likewise, we find that only large traders trade in the direction of the AF earnings surprises after earnings announcements. This evidence is consistent with distinct sets of traders explaining, at least in part, the two post-earnings-announcement drifts. Specifically, it suggests that small traders create upward (downward) price pressure on stocks in the post-announcement period for firms with good (bad) RW news. Likewise, it suggests that large traders create upward (downward) price pressure on stocks in the post-announcement period for firms with good (bad) AF news.

If the RW and AF drifts are manifested as an under-reaction to earnings news by small and large traders, respectively, then the more these traders act on the earnings news during the announcement period, the lower the subsequent drift should be. Consistent with this expectation, we find that when small traders trade more intensely in the direction of the RW earnings surprises *during the announcement period*, the magnitude of the subsequent RW drift is lower. Likewise, when large traders trade more intensely in the direction of the AF earnings surprises *during the announcement period*, the magnitude of the AF drift is lower.

Because large traders are considered *relatively* more sophisticated than are small traders (e.g., as evidenced by their more sophisticated earnings expectation model), the nature of the under-reaction by small and large traders and the related drifts is likely qualitatively different. We conduct several analyses to further explore and contrast the RW and AF drifts. First, we expect that the timing of post-announcement trades is a function of investors' *relative* sophistication. If large traders are relatively more sophisticated than small traders, we expect that large traders end their trading on AF earnings surprises more quickly than small traders end their trading on RW earnings surprises. Consistent with our expectation, we find that large traders trade more at the early stage of the post-announcement period, whereas small traders spread their trades throughout the 60-day post-announcement period.

Second, we examine small and large trades around one- to four-quarter-ahead earnings announcements using the methodology employed by Bernard and Thomas (1990). Consistent with small traders being less sophisticated, we find that small trades around subsequent earnings announcements are predictable based on lagged one- to four-quarter RW earnings surprises. Consistent with large traders being *relatively* more sophisticated and their under-reaction being short-lived, we find no evidence of large trades at earnings announcements being associated with prior AF earnings surprises. This evidence suggests that the RW drift, but not AF drift, is largely explained by small or unsophisticated traders' failure to recognize the time-series property of earnings.

---

[2] Hirshleifer et al. (2008) use a proprietary sample of individual trades from a single discount brokerage, whereas we analyze a comprehensive sample of small and large trades using the New York Stock Exchange's Trade and Quote database. While our theory is not contingent on small traders being individual investors (i.e., our theory relies on small traders under-reacting to the less sophisticated, RW earnings surprises and large traders under-reacting to AF earnings surprises), we are able to reconcile our results to the results in Hirshleifer et al. (2008) using our small trade data. This evidence suggests that the individual investors in Hirshleifer et al. (2008) trade similarly to the small traders in this study's setting and that the differences in this study from Hirshleifer et al. (2008) stem more from the theories tested and research designs.

Third, we investigate whether the AF drift and post-announcement trades by large traders associated with the AF earnings surprise are more pronounced when the AF earnings surprise is more difficult to interpret. Prior research finds that the AF drift is greater when analysts' forecasts are more heterogeneous (Liang 2003), suggesting that investors have difficulty interpreting AF earnings surprises when analysts' forecasts are less correlated. Consistent with large traders requiring more post-earnings-announcement information to interpret earnings surprises when analysts' forecasts are more heterogeneous, we find that the AF drift and post-announcement trades by large traders associated with the AF earnings surprise are more pronounced when there is more heterogeneity among analysts' forecasts. This finding suggests that large traders' under-reaction and the related AF drift may be attributable, at least in part, to a longer price discovery process when earnings are more difficult to interpret.

Fourth, we investigate how post-earnings-announcement analyst forecast revisions affect large and small traders' post-announcement trading on earnings surprises. If large traders' under-reaction is corrected in the post-announcement period through subsequent price discovery, analyst forecast revisions may be one source of such information discovery (Gleason and Lee 2003).[3] We find that analyst forecast revisions are sluggish with respect to prior earnings information and that much of the post-announcement trading by large traders based on the AF earnings surprise is attributable to information in analyst forecast revisions that continue to be correlated with the AF earnings surprise well after the earnings announcement. This evidence further suggests a longer price discovery explanation for the AF drift with analyst forecast revisions serving as one source of information discovery.

Finally, we examine the magnitude of the RW and AF drifts and related post-announcement trading by small and large traders in more recent years. Results suggest that the magnitude of the AF drift decreases in more recent years (after 1999), while the RW drift does not change. This evidence is consistent with a more efficient price discovery process for information contained in the AF earnings surprise in more recent years as the AF drift may have received more attention. Given the relatively new discovery of and the little attention that the AF drift received during the earlier years in our sample, sophisticated investors may have not sufficiently exploited the AF drift during this period, which may explain the larger AF drift in earlier years.

This study makes the following contributions. First, we predict and find that small traders systematically trade in the direction of the RW earnings surprise after earnings announcements, whereas large traders' trading during the post-earnings-announcement period is in the direction of the AF earnings surprise. This evidence suggests that the RW and AF drifts are attributable, at least in part, to distinct sets of investors who use different earnings expectation models. Second, to our knowledge, this is the first study to link *announcement period trading* by distinct groups of investors to the *RW* and *AF drifts*. These findings are important in understanding the respective drifts and the investors who influence the size and type of the drifts. Third, we provide evidence that the two drifts are qualitatively different. The RW drift is largely explained by small traders' failure to understand the time-series property of earnings and therefore represents investor naiveté. In contrast, the AF drift may be explained by a longer price discovery process by large traders that highly correlates with subsequent analyst forecast revisions and increases with analysts' forecast heterogeneity. Finally, we find that the AF drift and related large trades during the post-announcement period have decreased in more recent years as the AF drift may have received more

[3] Gleason and Lee (2003) find drift associated with analyst forecast revisions, consistent with a price discovery process. Because the price discovery process is *systematically correlated* with the AF earnings surprise and the AF drift is mitigated when large trades react more strongly to the AF earnings surprise during the announcement period, we characterize large traders as under-reacting to the AF earnings surprise.

attention. In sum, we conclude that the two drifts are quite different despite both being manifested as an investor under-reaction to earnings news.

Section II develops hypotheses. Section III describes how we classify trades into buys and sells and how we partition small and large traders. Section IV presents our sample selection and descriptive statistics, and Section V presents the results of our hypotheses tests. Section VI provides evidence of additional differences between the AF and RW drifts. Section VII summarizes and concludes.

II. HYPOTHESIS DEVELOPMENT

Investors' Post-Announcement Trading

To date, there is no consensus regarding who or what causes either drift, random-walk-based or analyst-based. While prior studies have not investigated the source(s) of the AF drift (given its relatively new discovery), recent studies have begun to focus on how different investor groups influence the RW drift. Bartov et al. (2000) find that the RW drift decreases with the level of institutional ownership and conclude that institutional investors improve on the earnings-processing problems that cause the RW drift. Their finding implies that higher individual ownership exacerbates the RW drift. Similarly, Ke and Ramalingegowda (2005) find that quarterly changes in *transient* institutional ownership in the direction of RW earnings surprise (e.g., an increase in ownership after good news) are associated with lower RW drift. This evidence suggests that more aggressive trading by transient institutional investors at earnings announcements negates mispricing associated with RW earnings surprises.

More recently, Hirshleifer et al. (2008) use individual trading data to investigate whether naïve individual investors trade *contrary* to the RW earnings surprises at earnings announcements, which creates underpricing after good RW news and overpricing after bad RW news. Implicit in the "contrarian trading" hypothesis is the presence of a group of informed or more sophisticated traders who attempt to move prices based on the RW earnings surprises and a second group of naïve individual investors who trade contrary to the RW earnings surprises. Hirshleifer et al. (2008) use a proprietary sample of individual trades from a single discount brokerage firm and find no evidence that individual investors trade contrary to RW earnings surprises.[4]

To build on prior research, we investigate an alternative explanation for the RW drift (and AF drift). We hypothesize that the two distinct post-earnings-announcement drifts are attributable, at least in part, to distinct sets of investors (small and large traders) who use different earnings expectation models but fail to fully react to earnings news. Our expectations differ from Hirshleifer et al. (2008) in that our reasoning is based on two different investor groups (small and large traders) trading on two different earnings signals (RW and AF earnings surprises) with both investor groups under-reacting to their respective earnings signals, as opposed to two different investor groups trading on one signal (RW earnings surprises) with one naïve group trading contrary to the signal.

We base our hypothesis on four streams of prior research: (1) findings that during the *earnings-announcement period*, small traders react to RW earnings surprises, while large traders tend to react to more accurate AF earnings surprises,[5] (2) asset-pricing models that suggest that

---

4 Shanthikumar (2004) examines small and large trades with the NYSE Trades and Quotations data and finds inconclusive evidence regarding the cause of the post-announcement earnings drift. Given the relatively new discovery of the AF drift, Shanthikumar (2004) does not consider the joint effect of RW and AF earnings surprises on investor trades and their relation with the AF and RW drifts.

5 While prior research shows that analysts' quarterly forecasts are more accurate than earnings predicted by time-series models, Bradshaw et al. (2010) show that simple random-walk forecasts are more accurate than analyst forecasts over longer forecast horizons and for firms that are smaller, younger, or have limited analyst following. Analysts' superiority

equilibrium prices aggregate different information sets of all investors, (3) behavioral evidence that investors tend to under-react to news, and (4) evidence that arbitrage activities are often limited. Collectively, these factors create a sufficient condition to generate two distinct drifts.

Bhattacharya (2001) studies the trading volume of large and small traders during earnings-announcement periods and finds that small traders' earnings expectations resemble a seasonal random-walk. Battalio and Mendenhall (2005) employ order imbalance measures and find that around earnings announcements, small traders respond to RW earnings surprises while large traders respond to AF earnings surprises. Mikhail et al. (2007) find that, while large traders trade in the direction of analyst stock recommendations and earnings forecast revisions, small traders do not. Collectively, these findings suggest that large traders use a more sophisticated earnings expectation, consistent with large traders likely being wealthier and more informed than small traders (Easley and O'Hara 1987; Lee 1992). The implication of these findings is that different investor groups (i.e., small and large traders) systematically form different earnings expectations and trade accordingly.[6,7]

Prior equilibrium models suggest that investors with differing beliefs can impact prices (Grossman 1976; Grossman and Stiglitz 1980; Admati 1985; Kandel and Pearson 1995; Shleifer and Vishny 1997). These models suggest that, in equilibrium, stock prices reflect the weighted average beliefs of different investor groups, with the risk-bearing capacity of each investor group determining its relative weight. These models also suggest that investor groups with differing earnings expectations (e.g., small traders who form their earnings expectation and trade based on RW earnings surprises and large traders who form their earnings expectation and trade based on AF earnings surprises) can impact prices. Consistent with this implication, Walther (1997) finds that stock returns reflect both RW earnings surprises and AF earnings surprises.

Prior behavioral research has suggested that market participants generally underestimate the implications of current earnings for future earnings. For example, Maines and Hand (1996) and Calegari and Fargher (1997) find that both undergraduate business and M.B.A. students seem to underestimate the auto-regressive component of the seasonal changes in quarterly earnings in the process of forecasting earnings. Stevens and Williams (2004) find that undergraduate business students under-react to both positive and negative news. Likewise, Mendenhall (1991), Ali et al. (1992), Abarbanell and Bernard (1992), and Abarbanell and Bushee (1997) find that analysts underestimate the persistence of earnings surprises in revising their earnings forecasts. In addition, several studies (e.g., Bushee 2001; Ke and Petroni 2004; Bradshaw et al. 2004; Callen et al. 2005) find that institutional investors have difficulty interpreting financial information.[8]

Arbitrage activities can potentially reduce the magnitude of the drifts. However, prior research shows that arbitrage forces can be quite limited in financial markets (De Long et al. 1990a, 1990b; Shleifer and Vishny 1997). A maintained assumption of our analysis is that arbitrageurs do not

---

is also less prevalent when analysts forecast large changes in earnings per share. These findings suggest that under certain conditions, sophisticated (large) traders should focus on random-walk-based earnings surprises instead of analyst-based earnings surprises. If this is the case, then it would bias against finding results consistent with our expectation for differential trading for large and small traders.

6 Use of random-walk-based expectations by small traders is likely, in part, attributable to their having less access to analyst forecasts. Consistent with this explanation, we find in tests described later that small traders' reliance on analyst forecasts increases in more recent sample years as analyst forecasts are more accessible (e.g., on the Internet).

7 None of these prior studies link small or large trades to RW or AF earnings surprises in the *post-announcement* period, leaving the questions of which investors or trading patterns cause the respective drifts open.

8 Bushee (2001) provides evidence that some institutional investors over- (under-) weight short- (long-) term earnings in their valuation models. Likewise, Ke and Petroni (2004) find that even though sophisticated institutional investors anticipate the break of a string of consecutive earnings increases, their selling activities lag behind the decline of stock prices by one to two quarters. In addition, Bradshaw et al. (2004) and Callen et al. (2005) find that institutional investors suffer from biases in interpreting foreign earnings.

completely eliminate under-reaction, because if arbitrage activities quickly correct initial under-reaction to earnings news, these post-announcement drifts would not occur.

Overall, the implication of these four streams of research is that the RW and AF drifts could be explained by distinct groups of investors that form their earnings expectations differently, who both impact prices and systematically under-react to earnings news. Because small traders are more likely to use RW earnings models to form their earnings expectations, we expect that trading by small traders in the post-announcement period primarily contributes to the RW drift. Likewise, because large traders are more apt to base their earnings expectations on analyst forecasts, we anticipate that their trading contributes more to the AF drift if they likewise fail to fully react to earnings surprises. We formulate the following hypotheses related to the post-earnings-announcement trading behavior of small and large traders:

H1: Small traders are more likely to trade shares in the direction of seasonal random-walk-based earnings surprises after earnings announcements.

H2: Large traders are more likely to trade shares in the direction of analyst-based earnings surprises after earnings announcements.

Recent research questions whether small traders represent individuals or instead institutions that split their trades into smaller trades (Barclay and Warner 1993; Chakravarty 2001; Campbell et al. 2005), whereas other studies present evidence consistent with small traders representing individual investors (Easley and O'Hara 1987; Lee 1992; De Franco et al. 2007; Bhattacharya et al. 2007; Ayers et al. 2008). It is important to note that this study's hypotheses are not contingent on small traders being individual investors and large traders being institutions. Instead, our expectations rely solely on small traders under-reacting to the less sophisticated RW earnings surprises and large traders under-reacting to AF earnings surprises. To the extent that both small traders and large traders represent the same traders (e.g., institutions that split their trades into small trades but use analyst forecasts to create their earnings expectations), this should bias against finding evidence supporting our hypotheses that predict trading differences between small and large traders.

Effect of Earnings-Announcement Period Trading on Subsequent Drift

If small and large traders' activities help explain the RW and AF drifts, respectively, then we expect a relation between the two drifts and the respective trading intensity by small and large traders *around earnings announcements*. Specifically, if investors assimilate the earnings news more thoroughly and thus trade more intensely around earnings announcements on RW and AF earnings surprises, then the magnitude of the respective RW and AF drifts should be lower. Since we hypothesize that small traders contribute to the RW drift, we expect that the more they trade at the announcement period on RW earnings surprises, the lower the RW drift. Likewise, we expect that the more large traders trade at the announcement period on AF earnings surprises, the lower the AF drift:

H3: The more intense the earnings-announcement period trading by small traders on seasonal random-walk-based earnings surprises, the smaller the post-earnings-announcement drift associated with seasonal random-walk-based earnings surprises.

H4: The more intense the earnings-announcement period trading by large traders on analyst-based earnings surprises, the smaller the post-earnings-announcement drift associated with analyst-based earnings surprises.

III. BUY-SELL ORDER IMBALANCE FOR SMALL AND LARGE TRADERS

We obtain stock quotes and investor trade data from the New York Stock Exchange's Trade and Quote (TAQ) database for years 1993, the year the TAQ database begins, through 2005. Using

the Lee and Ready (1991) and Lee (1992) algorithm, we classify trades into buyer-initiated or seller-initiated trades. While the number of shares bought equals the number of shares sold in a transaction, the Lee (1992) algorithm identifies the likelihood that a transaction is buyer-initiated or seller-initiated.[9]

Specifically, we compare traded prices with quotes that are at least five seconds earlier. If the traded price is above the mid-point of the bid-ask spread, then we define the trade as a buy (*BUY*). If the traded price is below the mid-point of the bid-ask spread, then we define the trade as a sell (*SELL*). For day *k*, we add up all the buys $\Sigma_{n=1}^{M} BUY_{im}$ and all the sells $\Sigma_{n=1}^{N} SELL_{in}$, where *M* is the total number of buys on Day *k* and *N* is the total number of sells on Day *k*. We then compute daily buy-sell order imbalance (i.e., net-buy) BMS_i ($\Sigma_{m=1}^{M} BUY_{im} - \Sigma_{n=1}^{N} SELL_{in}$). A positive (negative) BMS_i indicates net-buying (net-selling) on Day *k*.

Because we are interested in *excess* trading activities, we subtract *normal* daily net-buy from the daily net-buy measure during the event period. We measure normal daily net-buy during the control period, a 40-day period from Day −45 to Day −6 relative to the earnings announcement Day *i*.[10] Specifically, the normal daily net-buy relative to Day *i* is the average daily net-buy computed as:

$$NBMS_i = \sum_{k \in [-45,-6]} BMS_{i,k}/T \tag{1}$$

where *T* is the total number of days with data available for *BMS*. Our daily excess net-buy measure is defined as $BMS_i - NBMS_i$.

Finally, to facilitate cross-sectional comparison of our daily excess net-buy measure, we scale it by the normal daily trading volume (i.e., daily trading volume during the control period). Daily trading volume, BPS_i, is defined as daily *BUY* plus daily *SELL* (i.e., $\Sigma_{m=1}^{M} BUY_{im} + \Sigma_{n=1}^{N} SELL_{in}$), and the normal daily trading volume relative to Day *i* is the average daily buy plus sell:

$$NBPS_i = \sum_{k \in [-45,-6]} BPS_{i,k}/T \tag{2}$$

where *T* is the total number of days with data available for *BPS*.[11] The excess net-buy on Day *k* during the event period (relative to the earnings announcement Day *i*) is therefore defined as net-buy in excess of the control period net-buy, scaled by the control period volume:

---

[9] Based on Lee (1992), a transaction is classified as buyer- (seller-) initiated if a buyer (seller) demands immediate execution (i.e., a market order). Generally there are three types of orders: market orders, limit orders, and standing orders. A market order demands immediate execution. A limit order is an order to be executed when a pre-specified price is attained. A standing order is an order to buy or sell shares at the best available price over a certain period of time, during which a broker has to use discretion. After the opening trade, a trade occurs only when a market order arrives. If a market order to buy is filled by a limit order to sell, the trade is classified as a buyer-initiated trade. If a market order to sell is filled by a limit order to buy, the trade is classified as a seller-initiated trade. Sometimes, the size of a market order and the size of a limit order are not equal. If one large market order to buy (sell) is filled by several small limit orders to sell (buy) (and possibly partially filled by the specialist), then the trade is classified as one large buyer- (seller-) initiated trade. If several small market orders to buy (sell) are filled by one large limit order to sell (buy) (and possibly partially filled by the specialist), then the trades are classified as several buyer- (seller-) initiated trades. Lee and Radhakrishna (2000), using the Trades, Orders, Reports, and Quotes database that contains information on trade directions and trader identities, find that while few (6 percent) of the total market orders are split up in execution, a much larger portion (24 percent) of the total market orders is batched in execution. However, despite the prevalence of order batching, trade size is still highly effective in separating large traders and small traders.

[10] Results are similar if we do not base our test on *excess* net-buys—i.e., if we do not create our measure of abnormal daily net-buys by subtracting the normal daily net-buys during the control period.

[11] We require *T* to be at least 20 to compute $NBPS_i$ and $NBMS_i$.

$$EXBMS_{i,k} = \frac{BMS_{i,k} - NBMS_i}{NBPS_i}. \tag{3}$$

This measure captures the excess net-buying activities on Day k. Once we have a measure of daily excess net-buy, we calculate the average excess net-buy during an event window $[k_1, k_2]$ (relative to the earnings announcement Day i) as $\Sigma_{k=k_1}^{k_2} EXBMS_{i,k} / (k_2 - k_1 + 1)$.

To analyze the trading behavior of small and large traders, we compute excess net-buy for small trades and large trades separately.[12] Specifically, if the dollar value of a round lot trade is below \$5,000, then we classify the trade as a small trade. If the dollar value of a round lot trade is above \$30,000, then we classify the trade as a large trade. Note that there are no exact cut-offs for defining small and large trades. The \$5,000 and \$30,000 cut-offs are used here to distinguish small and large traders that prior research suggests form earnings expectations based on RW and AF earnings surprises, respectively. We use a buffer of \$25,000 between the \$5,000 cut-off for small trades and the \$30,000 cut-off for large trades to reduce the ambiguity that a trade is initiated by a small or large trader (Lee and Radhakrishna 2000).[13]

IV. SAMPLE AND DESCRIPTIVE STATISTICS

Sample Selection

We obtain actual quarterly earnings, analyst forecasts of quarterly earnings, and earnings announcement dates from the I/B/E/S unadjusted file for the period 1993–2005.[14] We also require firms to have I/B/E/S earnings for the same quarter in the prior year to calculate RW earnings surprise. We define RW earnings surprise (*RW*) as:

$$RW = (EPS_t - EPS_{t-4})/P_{t-1} \tag{4}$$

where EPS_t is the I/B/E/S actual earnings per share (EPS) for quarter t, EPS_{t-4} is the I/B/E/S actual EPS for quarter t–4, and P_{t-1} is stock price at the beginning of quarter t.[15,16]

Similar to Equation (4), we define AF earnings surprise (*AF*) as the I/B/E/S actual EPS for quarter t minus the analyst forecast for quarter t, deflated by beginning of quarter t stock price in Compustat. We use the single-most recent forecast made by the timeliest analyst(s) prior to the earnings announcement date as our analyst forecast.[17] When there is more than one analyst forecast on the most recent day, we take the mean of these forecasts. Consistent with prior studies, we convert *RW* and *AF* into earnings surprise deciles based on the magnitude of earnings surprises in

---

[12] Battalio and Mendenhall (2005) classify trade size by number of shares. We classify traders based on the dollar amount involved, because this is a relatively more direct measure of investor wealth (Bhattacharya 2001; Bhattacharya et al. 2007).

[13] As a sensitivity analysis, we use alternative cut-offs such as \$10,000 and \$15,000 for small trades and \$20,000 and \$25,000 for large trades. Results are robust to these alternative cut-offs.

[14] We use earnings announcements dates from I/B/E/S because we require analyst forecasts issued prior to I/B/E/S earnings announcement dates. The earnings announcement dates in I/B/E/S differ from Compustat by no more than one calendar day.

[15] While prior research uses Compustat actual earnings to calculate RW earnings surprises, we use I/B/E/S actual earnings to be consistent with the definition of AF earnings surprises. Livnat and Mendenhall (2006) demonstrate that using I/B/E/S actual earnings does not significantly alter the RW drift. Our results are similar when we use Compustat actual earnings to calculate RW earnings surprises.

[16] Following Livnat and Mendenhall (2006), we define *RW* and *AF* so that they only differ in expected earnings (i.e., they have the same reported actual earnings and same deflator). In sensitivity analyses, we use the standard deviation of prior RW surprises over the past eight quarters as the deflator for RW (Bernard and Thomas 1990) and obtain similar results.

[17] Brown and Caylor (2005) argue that the most recent forecast is a better proxy for expected earnings than the consensus forecast for two reasons. First, using the most recent forecast mitigates the effects of pre-announcements. Second, evidence suggests that the most recent forecast is more closely related to the stock price reaction to earnings announcements (Brown and Kim 1991). When we use the consensus analyst forecast as the benchmark for our analysis, we obtain qualitatively similar results.

the population in the same quarter. Similar to Livnat and Mendenhall (2006), we focus on a 60-day post-earnings-announcement period and require firms to have stock return data from the CRSP Daily Stock file during the period [−1, +65] surrounding earnings announcement dates. We obtain 129,215 observations based on these selection criteria. Finally, we merge this data with excess net-buy measures and require no missing excess net-buy measures for both the [−1, +1] and [+6, +65] periods following earnings announcements.[18] For regression analyses, we delete observations in the top or bottom 1 percentile in the distributions of excess net-buy measures. Our final sample includes 73,469 observations for 5,661 firms.[19]

Descriptive Statistics for the Earnings Announcements Period

Table 1 displays abnormal returns and average excess net-buy for small and large traders during the three-day [−1, +1] earnings-announcement period, for quintiles of RW and AF earnings surprises. Panel A reports the magnitude of the average value of three-day cumulative abnormal returns for the 25 portfolios formed by the RW and AF earnings surprises quintiles. Within each AF quintile, abnormal returns generally increase in RW quintiles, and within each RW quintile, abnormal returns generally increase in AF quintiles, consistent with Walther (1997).

Panel B of Table 1 shows the mean values of three-day average excess net-buy for small traders around the earnings-announcement period for the 25 portfolios formed by RW and AF quintiles. Within each AF quintile, small traders' excess net-buy generally increases in RW quintiles. On the other hand, within each RW quintile, the pattern with respect to AF quintiles is less clear. Consistent with Bhattacharya (2001) and Battalio and Mendenhall (2005), our data suggest that small traders react more to RW earnings surprises during the announcement period.

Panel C of Table 1 shows the mean value of three-day average excess net-buy for large traders around the earnings-announcement period. Within each RW quintile, large traders' excess net-buy generally increases in AF quintiles. However, within each AF quintile, we do not observe any systematic pattern between large traders' excess net-buy and RW quintiles. In sum, data indicate that large traders' earnings-announcement period trading responds only to AF earnings surprises, consistent with Battalio and Mendenhall (2005).[20]

Descriptive Statistics for the Post-Earnings-Announcement Period

Table 2 provides statistics on the 60-day [+6, +65] cumulative abnormal returns during the post-earnings-announcement period, and small and large trader excess net-buy for RW and AF quintiles. Panel A reports the average 60-day cumulative abnormal returns for the 25 portfolios sorted on RW and AF earnings surprises. Overall, the average abnormal returns during the post-earnings-announcement period generally increase both in AF and RW earnings surprises, although the results are not monotonic within quintiles. Regarding the magnitudes of the two drifts, we find that the spread in abnormal returns between *AF*5 and *AF*1 is 4.8 percent, while the spread between

[18] Because our focus is on trading during the *post*-earnings-announcement period, we do not analyze trading data during the [+2, +5] window to avoid ambiguity as prior research (Morse 1981; Bamber 1987) finds that volume reactions extend up to day +5.

[19] To determine if results are robust to less stringent sample criteria, we relax the requirement that each firm-quarter has 60 days of post-announcement trading data. Specifically, we compute the daily average post-announcement buy-sell order imbalance for firm-quarters that have at least 20 days of buy-sell order imbalance data during the [+6, +65] post-announcement period. Relaxing this requirement increases the total number of observations in our main analysis to 94,917 (out of 112,876 observations after we require TAQ data for the 40-day control period and 3-day announcement period) and yields similar results.

[20] We note that the contrarian trading hypothesis in Hirshleifer et al. (2008) predicts that small traders trade contrary to RW earnings surprises (not AF earnings surprise) during the announcement period. We find no evidence of contrarian trading by either small or large traders during the announcement period. Thus, like Hirshleifer et al. (2008), our evidence does not support the contrarian trading hypothesis.

TABLE 1

Abnormal Returns and Excess Net-Buy for Small and Large Traders during [−1, +1] Earnings-Announcement Period

Panel A: Cumulative Abnormal Returns during [−1, +1]

	*RW*1	*RW*2	*RW*3	*RW*4	*RW*5	Average
*AF*1	−0.030	−0.027	−0.023	−0.017	−0.017	−0.026
*AF*2	−0.006	−0.011	−0.009	−0.010	−0.003	−0.009
*AF*3	−0.002	0.005	0.008	0.008	0.001	0.006
*AF*4	0.006	0.013	0.020	0.022	0.015	0.017
*AF*5	0.017	0.020	0.021	0.030	0.032	0.027
Average	−0.010	−0.004	0.003	0.011	0.015	

Panel B: Average Excess Net-Buy for Small Trades during [−1, +1]

	*RW*1	*RW*2	*RW*3	*RW*4	*RW*5	Average
*AF*1	0.008	0.033	0.044	0.027	0.062	0.026
*AF*2	0.017	0.034	0.049	0.049	0.052	0.040
*AF*3	0.038	0.033	0.057	0.075	0.061	0.055
*AF*4	0.021	0.034	0.045	0.053	0.082	0.050
*AF*5	0.031	0.058	0.043	0.050	0.082	0.059
Average	0.018	0.034	0.050	0.056	0.074	

Panel C: Average Excess Net-Buy for Large Trades during [−1, +1]

	*RW*1	*RW*2	*RW*3	*RW*4	*RW*5	Average
*AF*1	−0.012	−0.029	−0.012	0.001	−0.031	−0.018
*AF*2	−0.005	0.006	0.000	−0.001	0.051	0.005
*AF*3	0.041	0.024	0.041	0.034	0.029	0.034
*AF*4	0.050	0.044	0.078	0.039	0.037	0.049
*AF*5	0.069	0.061	0.057	0.098	0.081	0.078
Average	0.017	0.015	0.032	0.037	0.050	

Panel A shows cumulative abnormal return during [−1, +1] window around earnings announcements for seasonal random-walk earnings surprise quintiles (*RW*1–5) and for analyst-based earnings surprise quintiles (*AF*1–5). Abnormal return during the earnings-announcement period is defined as the firm return in excess of the corresponding Fama-French size and book-to-market six-portfolio benchmark return. Seasonal random-walk earnings surprise is defined as I/B/E/S actual EPS for quarter t minus I/B/E/S actual EPS for quarter t–4, deflated by price per share at the beginning of quarter t. Analyst-based earnings surprise is defined as I/B/E/S actual EPS for quarter t minus most recent forecast for quarter t, deflated by price per share at the beginning of quarter t.

Panel B shows average excess net-buy for small trades during [−1, +1] window around earnings announcements for seasonal random-walk earnings surprise quintiles (*RW*1–5) and for analyst-based earnings surprise quintiles (*AF*1–5). Small trade is defined as a trade at maximum $5,000.

Panel C shows average excess net-buy for large trades during [−1, +1] window around earnings announcements for seasonal random-walk earnings surprise quintiles (*RW*1–5) and for analyst-based earnings surprise quintiles (*AF*1–5). Large trade is defined as a trade at minimum $30,000.

*RW*5 and *RW*1 is 4.3 percent. The spreads are not statistically different (p = 0.60), but they are comparable to those reported by Livnat and Mendenhall (2006), who find a spread of 4.9 percent between the top and the bottom AF deciles, and 3.8 percent between the top and the bottom RW earnings surprise deciles in their sample.

TABLE 2

Abnormal Returns and Excess Net-Buy for Small and Large Trades during [+6, +65] Post-Earnings-Announcement Period

Panel A: Cumulative Abnormal Returns during [+6, +65]

	***RW*1**	***RW*2**	***RW*3**	***RW*4**	***RW*5**	**Average**
*AF*1	−0.030	−0.017	−0.005	−0.001	−0.015	−0.023
*AF*2	−0.030	−0.033	−0.026	−0.010	−0.020	−0.025
*AF*3	−0.022	−0.019	−0.004	−0.008	0.011	−0.008
*AF*4	−0.009	−0.023	0.004	0.024	0.012	0.005
*AF*5	0.004	−0.020	0.013	0.012	0.053	0.025
Average	−0.021	−0.024	−0.008	0.005	0.022	

Panel B: Average Excess Net-Buy for Small Trades during [+6, +65]

	***RW*1**	***RW*2**	***RW*3**	***RW*4**	***RW*5**	**Average**
*AF*1	0.006	0.013	0.020	0.021	0.025	0.012
*AF*2	0.010	0.012	0.023	0.024	0.018	0.017
*AF*3	−0.008	0.004	0.020	0.028	0.035	0.018
*AF*4	−0.002	0.005	0.011	0.023	0.030	0.016
*AF*5	0.007	0.011	0.008	0.016	0.029	0.019
Average	0.005	0.009	0.018	0.023	0.028	

Panel C: Average Excess Net-Buy for Large Trades during [+6, +65]

	***RW*1**	***RW*2**	***RW*3**	***RW*4**	***RW*5**	**Average**
*AF*1	0.025	0.002	−0.009	0.001	−0.002	0.011
*AF*2	0.020	0.005	0.004	−0.005	−0.003	0.004
*AF*3	0.022	0.013	0.007	0.007	−0.006	0.008
*AF*4	0.025	0.014	0.018	0.012	0.004	0.013
*AF*5	0.013	0.016	0.025	0.019	0.014	0.015
Average	0.022	0.008	0.008	0.007	0.006	

Panel A shows cumulative abnormal return during [+6, +65] window following earnings announcements for seasonal random-walk earnings surprise quintiles (*RW*1–5) and for analyst-based earnings surprise quintiles (*AF*1–5). Abnormal return during the post-earnings-announcement period is defined as the firm return in excess of the corresponding Fama-French size and book-to-market six-portfolio benchmark return. Seasonal random-walk earnings surprise is defined as I/B/E/S actual EPS for quarter *t* minus I/B/E/S actual EPS for quarter *t*–4, deflated by price per share at the beginning of quarter *t*. Analyst-based earnings surprise is defined as I/B/E/S actual EPS for quarter *t* minus most recent forecast for quarter *t*, deflated by price per share at the beginning of quarter *t*.

Panel B shows average excess net-buy for small trades during [+6, +65] window following earnings announcements for seasonal random-walk earnings surprise quintiles (*RW*1–5) and for analyst-based earnings surprise quintiles (*AF*1–5). Small trade is defined as a trade at maximum $5,000.

Panel C shows average excess net-buy for large trades during [+6, +65] window following earnings announcements for seasonal random-walk earnings surprise quintiles (*RW*1–5) and for analyst-based earnings surprise quintiles (*AF*1–5). Large trade is defined as a trade at minimum $30,000.

Panel B of Table 2 presents the mean value of average excess net-buy for small trades for the 60-day post-earnings-announcement period. Results suggest that, within each AF quintile, small trader excess net-buy generally increases in RW quintiles. Thus, consistent with H1, small traders continue to trade on information in RW earnings surprises in the post-announcement period. We

also find that small traders' excess net-buys decline in AF quintiles in the second and third RW quintiles. Therefore, it appears that, while small traders continue to trade in the direction of RW earnings surprises, they also trade in the opposite direction of AF earnings surprises.

Panel C of Table 2 focuses on the average excess net-buy for large traders during the 60-day post-earnings-announcement period. In the second, third, fourth and fifth quintiles of the RW earnings surprises, large traders' excess net-buys generally increase in AF quintiles. Thus, consistent with H2, large traders continue to trade on information in AF earnings surprises in the post-announcement period. In the first, second, third, and fourth AF quintiles, large traders' excess net-buys generally decrease in RW quintiles. Therefore, it appears that, while large traders continue to trade in the direction of AF earnings surprises, they also trade in the opposite direction of RW earnings surprises.

V. RESULTS OF HYPOTHESES TESTS

Trading during the Earnings-Announcement Period

We first examine trading during the earnings-announcement period. Specifically, to estimate Equation (5), we regress cumulative abnormal returns (*CAR*) or average excess net-buy (*EXBMS*) for small or large traders during the three-day (i.e., $[-1, +1]$) earnings-announcement period on the decile ranks of the seasonal random-walk-based earnings surprises (*RW*) and the decile ranks of the analyst-based earnings surprises (*AF*). Following Bernard and Thomas (1989), we scale *RW* and *AF* to be between 0 and 1 in all regression analyses:

$$CAR_{it}/DRIFT_{it}/EXBMS\_Small_{it}/EXBMS\_Large_{it} = \beta_0 + \beta_1 RW_{it} + \beta_2 AF_{it} + \varepsilon_{1it}. \quad (5)$$

We report regression results estimating Equation (5) in Panel A of Table 3. All t-statistics in this table and subsequent tables are adjusted for clustering by firm (Rogers 1993). When the three-day *CAR* is the dependent variable, the coefficients on the seasonal random-walk earnings surprise, *RW*, and the analyst-based earnings surprise, *AF*, are positive and significant ($\beta_1 = 0.011$, t $= 10.59$, $p < 0.01$; $\beta_2 = 0.056$, t $= 46.46$, $p < 0.01$).[21] An F-test indicates that β_2 is significantly greater than β_1 ($p < 0.01$), suggesting that, while the market as a whole reacts to both the RW earnings surprises and the AF earnings surprises (Walther 1997), it reacts more intensely to AF earnings surprises.[22]

---

[21] *RW* and *AF* are not highly correlated (correlation = 0.34), which suggests that the "earnings news" conveyed by the same reported number can generate significant differential implications if investors have different earnings expectation models. Similarly, regression diagnostics do not suggest that multicollinearity is an issue. The variance inflation factor (VIF) for Equation (5) is 1.13, well below the conventional value of 10, indicating a multicollinearity problem. In supplemental analysis, we analyze small and large trades for firms with no I/B/E/S analyst following (i.e., limited or no analyst following). We find that during the earnings-announcement period small traders react to RW earnings surprises, while large trades are *not* associated with RW earnings surprises. We also find that after earnings announcements, small trades are positively associated with RW earnings surprises, while large trades are negatively associated with RW earnings surprises. These results indicate that small traders react to RW earnings surprises at the earnings announcements and continue to trade in the direction of RW earnings surprises after the earnings announcements. Large traders, however, are more sophisticated, potentially understand that RW earnings surprises are inferior signals, and update their expectations based on other sources of information, such as market, firm, and industry news events. Within the sample of firms without I/B/E/S analyst following, these results also hold for firms in the top tercile of market value and top tercile of large trader trading volume, which suggests that transaction costs do not explain large traders' lack of trading on RW earnings surprises.

[22] In sensitivity analyses, we analyze those firms with conflicting signs of RW and AF earnings surprises. We find that 18.8 percent of the time a firm meets/beats analyst forecasts but fails to beat earnings of the same quarter in the prior year, and 12.5 percent of the time a firm meets/beats prior earnings but reports earnings lower than analysts' expectations. We find that firms with conflicting signs of RW and AF earnings surprises are not significantly different in size, trading volume, expected growth, number of analyst following, or analysts' forecasts dispersion. Finally, we find that the RW and AF drifts and related trading patterns are not significantly affected by whether the sign of RW and AF earnings surprises agree or disagree.

TABLE 3

Regressions of Abnormal Returns and Excess Net-Buy for Small and Large Trades

Panel A: Three-Day Announcement Period

	CAR	Small Trades	Large Trades
Intercept	−0.031	0.011	−0.026
	(−43.31)***	(4.08)***	(−4.38)***
RW	0.011	0.055	0.005
	(10.59)***	(13.84)***	(0.55)
AF	0.056	0.016	0.106
	(46.46)***	(3.81)***	(11.41)***
R^2 (%)	5.78	0.38	0.26
n	73,469	73,469	73,468

Panel B: 60-Day Post-Announcement Period

	DRIFT	Small Trades	Large Trades
Intercept	−0.047	0.004	0.012
	(−13.29)***	(3.46)***	(5.84)***
RW	0.040	0.030	−0.018
	(6.73)***	(15.21)***	(−5.77)***
AF	0.043	−0.005	0.015
	(7.47)***	(−2.58)***	(4.30)***
R^2 (%)	0.26	0.28	0.16
n	73,469	73,469	73,469

*, **, *** Represent significance levels at less than two-tailed 10 percent, 5 percent, and 1 percent, respectively.
All t-statistics in parentheses are adjusted for clustering by firm (Rogers 1993).

Table 3 shows regression of cumulative abnormal return, average excess net-buy of small trades, and average excess net-buy of large trades during [−1, +1] earnings announcement window and during [+6, +65] window following earnings announcements. Abnormal return during the post-earnings-announcement period is defined as the firm return in excess of the corresponding Fama-French size and book-to-market six-portfolio benchmark return. *RW* is the decile rank of seasonal random-walk earnings surprise converted to [0, 1], and *AF* is the decile rank of analyst-based earnings surprise converted to [0, 1]. Seasonal random-walk earnings surprise is defined as I/B/E/S actual EPS for quarter *t* minus I/B/E/S actual EPS for quarter *t*–4, deflated by price per share at the beginning of quarter *t*. Analyst-based earnings surprise is defined as I/B/E/S actual EPS for quarter *t* minus most recent forecast for quarter *t*, deflated by price per share at the beginning of quarter *t*. Small trade is defined as a trade at maximum $5,000. Large trade is defined as a trade at minimum $30,000.

When the small traders' excess net-buy (*EXBMS_Small*) is the dependent variable, the coefficients on *RW* and *AF* are positive and significant ($\beta_1 = 0.055$, t = 13.84, $p < 0.01$; $\beta_2 = 0.016$, t = 3.81, $p < 0.01$), but the coefficient on AF is much smaller ($p < 0.01$) than that on *RW*. Thus, small traders react more to the RW earnings surprises than to the AF earnings surprises.

When the large traders' excess net-buy (*EXBMS_Large*) is the dependent variable, the coefficient on *RW* is insignificant ($\beta_1 = 0.005$, t = 0.55, p = 0.58) and the coefficient on *AF* is positive and significant ($\beta_2 = 0.106$, t = 11.41, $p < 0.01$). Thus, large traders appear to respond only to analyst-based earnings surprises during the announcement period, consistent with Battalio and Mendenhall's (2005) result.

Testing H1 and H2: Trading during the Post-Earnings-Announcement Period

Panel B of Table 3 provides the results of estimating Equation (5) for the 60-day (i.e., [+6, +65]) period *after* the earnings announcements. When the cumulative abnormal return, *DRIFT*

(i.e., the post-earnings-announcement drift), is the dependent variable, the coefficients on *RW* and *AF* are positive and significant ($\beta_1 = 0.040$, t = 6.73, p < 0.01; $\beta_2 = 0.043$, t = 7.47, p < 0.01), consistent with the existence of distinct RW and AF drifts (Livnat and Mendenhall 2006). The coefficients for *RW* and *AF* are not statistically different (p = 0.71).

Panel B of Table 3 also presents tests of H1 and H2. Consistent with H1, when we use the post-earnings-announcement small trader excess net-buy as the dependent variable, the coefficient on *RW* is positive and significant ($\beta_1 = 0.030$, t = 15.21, p < 0.01). This evidence indicates that small traders continue to trade in the direction of RW earnings surprises after earnings announcements. Consistent with H2, when the post-earnings-announcement large trader excess net-buy is the dependent variable, the coefficient on *AF* is positive and significant ($\beta_2 = 0.015$, t = 4.30, p < 0.01). This result indicates that large traders continue to trade in the direction of analyst-based earnings surprises after earnings announcements. In combination, this evidence is consistent with distinct sets of traders explaining, at least in part, the two post-earnings-announcement drifts.

Results in Table 3, Panel B also indicate a negative and significant coefficient on *AF* (−0.005, t = −2.58, p = 0.01) in the small trader net-buy regression, suggesting that, while small traders continue to trade in the direction of the RW earnings surprises, they also trade in the opposite direction of the AF earnings surprises. Similarly, we find a negative and significant coefficient on *RW* (−0.018, t = −5.77, p < 0.01) in the large trader net-buy regression, suggesting that, while large traders continue to trade in the direction of the AF earnings surprises after the earnings announcements, they also trade in the opposite direction of the RW earnings surprises. These results are similar to the relations we find in univariate comparisons.

A plausible explanation for large traders trading in the opposite direction of RW earnings surprises during the post-announcement period is that they provide liquidity to small traders trading on RW earnings surprises. Specifically, because large traders have an earnings expectation model different from that of small traders and consequently a different and independent estimate of the intrinsic value of the same firm, they may sell a security when the demand by small traders for the security pushes its price above large traders' private valuation of the security.

Liquidity trading is less likely to explain the negative association between trades by less sophisticated small traders and AF earnings surprises and, thus, we investigate an alternative explanation: attention-based trading. Lee (1992), Battalio and Mendenhall (2005), and Hirshleifer et al. (2008) find that small traders are net buyers irrespective of the direction of the earnings surprise. Barber and Odean (2008) argue that retail investors are heavily influenced by a "news attention" effect that prompts buying whenever a company is in the news. To test the possibility that the negative coefficient on *AF* may be due primarily to net-buying by small traders when *AF* is negative, we add to Equation (5) an indicator variable, *NegAF*, which equals 1 if the analyst-based earnings surprise is negative, and 0 otherwise. Consistent with individual net-buys following negative AF surprises explaining the negative association between small trades and AF surprises, the coefficient on *NegAF* is positive and significant, while the coefficient on *AF* becomes insignificant.[23] Further, following Barber and Odean (2008), we add trading volume, *Vol*, to Equation (5) to proxy for news attention and interact it with *NegAF*. We define *Vol* as the average daily trading volume during the earnings-announcement period [−1, +1], scaled by the average daily trading volume during the control period [−45, −6]. Consistent with attention-based trading, the coefficient on *NegAF* × *Vol* is positive and significant while the coefficients on *NegAF* and *AF* are insignificant.[24]

---

[23] To assure that attention trading does not explain the negative coefficient for *RW* in the large trade regression, we run a similar analysis for large trades, including an indicator variable, *NegRW*, which equals 1 if the seasonal random-walk earnings surprise is negative, and 0 otherwise. Including *NegRW* does not affect the negative coefficient for *RW*.

[24] The coefficient on *Vol* is also positive and significant. The positive and significant coefficients on *Vol* and *NegAF* × *Vol*

Testing H3 and H4: Effects of Announcement Period Trading Intensity on the Post-Earnings-Announcement Drift

H3 (H4) predicts that the intensity of earnings-announcement period trading by small (large) traders on RW (AF) earnings surprises reduces the magnitude of the post-earnings-announcement drift associated with RW (AF) earnings surprises. We run the following regression to test H3 and H4:

$$\begin{aligned} DRIFT_{it} = {} & \theta_0 + \theta_1 RW_{it} + \theta_2 AF_{it} + \theta_3 EXBMS\_Small_{it} + \theta_4 EXBMS\_Large_{it} \\ & + \theta_5 RW_{it} \cdot EXBMS\_Small_{it} + \theta_6 AF_{it} \cdot EXBMS\_Small_{it} + \theta_7 RW_{it} \cdot EXBMS\_Large_{it} \\ & + \theta_8 AF_{it} \cdot EXBMS\_Large_{it} + \theta_9 RW_{it} \cdot TransCost_{it} + \theta_{10} AF_{it} \cdot TransCost_{it} \\ & + \theta_{11} TransCost_{it} + \varepsilon_{2it}, \end{aligned} \tag{6}$$

where *DRIFT* is the 60-day [+6, +65] cumulative abnormal return. *EXBMS_Small* and *EXBMS_Large* are the three-day [−1, +1] announcement period excess net-buy for small and large traders, respectively. *TransCost* is a measure of transaction cost that takes on values between 0 and −1 and is defined as −1 times the average of scores on the following three dimensions: (1) the decile ranking (scaled to between 0 and 1) of market value at the end of the earnings announcement quarter, (2) the decile ranking (scaled to between 0 and 1) of trading volume over the preceding fiscal year ending in the earnings announcement quarter, and (3) an indicator variable that equals 1 if price at the end of the earnings announcement quarter is greater than \$10, and 0 otherwise (Bhushan 1994; Kimbrough 2005).[25]

We test whether small (large) trades during the announcement period attenuate the relation between RW (AF) earnings surprises and post-announcement returns by interacting small (large) trades with RW (AF) earnings surprises in Equation (6). In this regression, the coefficient on the interaction term between small trades and RW earnings surprise (*RW·EXBMS_Small*) captures whether earnings announcement trading by small traders attenuates the association between RW earnings surprises and post-announcement returns (i.e., the RW drift), as posited by H3. The coefficient on the main effect of announcement period small trades (*EXBMS_Small*) captures the overall relation between small trades and post-announcement returns, unconditioned on RW earnings surprises. We expect a negative coefficient on *RW·EXBMS_Small.*

Likewise, the coefficient on the interaction term between large trades and AF earnings surprises (*AF·EXBMS_Large*) captures whether earnings announcement trading by large traders attenuates the association between the AF earnings surprises and post-announcement returns (i.e., the AF drift), as posited by H4. Similarly, the coefficient on the main effect of large trades (*EXBMS_Large*) captures the overall relation between announcement period large trades and post-announcement returns, unconditioned on AF earnings surprises. We expect a negative coefficient on *AF·EXBMS_Large*.

Transaction costs reduce the incentives for arbitrage activities and, *ceteris paribus*, should enhance the post-earnings-announcement drift. Accordingly, we expect the coefficients on *RW·TransCost* and *AF·TransCost* to be positive to the extent that transaction costs limit arbitrage opportunities associated with the RW and AF drifts.

Table 4 presents the results from estimating Equation (6). In Column (1), we report the results from estimating Equation (6) separately for small traders (i.e., excluding the $EXBMS\_Large_{it}$

imply asymmetric net buying by small traders in response to the news effect, with small traders reacting more to negative AF news. This intriguing finding is consistent with negative news drawing more attention-based trading by small traders and warrants further study by future research.

[25] Results are similar if we use bid-ask spread as an alternative measure of transaction cost.

TABLE 4

Effects of Excess Net-Buy during Earnings Announcements Period on Post-Earnings-Announcement Drift

	(1)	(2)	(3)
Intercept	−0.066	−0.067	−0.066
	(−4.08)***	(−4.14)***	(−4.07)***
RW	0.097	0.097	0.097
	(4.07)***	(4.06)***	(4.07)***
AF	0.075	0.077	0.075
	(3.13)***	(3.23)***	(3.13)***
RW·EXBMS_Small	−0.044	—	−0.046
	(−3.20)***		(−3.29)***
AF·EXBMS_Small	0.009	—	0.011
	(0.47)		(0.56)
RW·EXBMS_Large	—	0.015	0.016
		(1.72)*	(1.82)*
AF·EXBMS_Large	—	−0.017	−0.018
		(−2.07)**	(−2.16)**
RW·TransCost	0.076	0.077	0.076
	(2.67)***	(2.71)***	(2.68)***
AF·TransCost	0.042	0.044	0.042
	(1.49)	(1.54)	(1.47)
EXBMS_Small	0.042	—	0.042
	(3.60)***		(3.61)***
EXBMS_Large	—	0.001	0.001
		(0.24)	(0.10)
TransCost	−0.027	−0.029	−0.027
	(−1.42)	(−1.54)	(−1.41)
R^2 (%)	0.36	0.33	0.37
Obs.	73,469	73,469	73,469

*, **, *** Represent significance levels at less than two-tailed 10 percent, 5 percent, and 1 percent, respectively.

All t-statistics in parentheses are adjusted for clustering by firm (Rogers 1993).

Table 4 shows regression of cumulative abnormal return during [+6, +65] window following earnings announcements. Abnormal return during the post-earnings-announcement period is defined as the firm return in excess of the corresponding Fama-French size and book-to-market six-portfolio benchmark return. *RW* is the decile rank of seasonal random-walk earnings surprise converted to [0, 1], and *AF* is the decile rank of analyst-based earnings surprise converted to [0, 1]. Seasonal random-walk earnings surprise is defined as I/B/E/S actual EPS for quarter *t* minus I/B/E/S actual EPS for quarter *t*–4, deflated by price per share at the beginning of quarter *t*. Analyst-based earnings surprise is defined as I/B/E/S actual EPS for quarter *t* minus most recent forecast for quarter *t*, deflated by price per share at the beginning of quarter *t*. *EXBMS_Small* is the average excess net-buy for small trades during [−1, +1] window around earnings announcements. Small trade is defined as a trade at maximum $5,000. *EXBMS_Large* is the average excess net-buy for large trades during [-1, +1] window around earnings announcements. Large trade is defined as a trade at minimum $30,000. *TransCost* is transaction cost taking on values between 0 and −1, and defined as −1 times the average of score on the following three dimensions: (1) the decile ranking of market value at the end of earnings announcement quarter, (2) the decile ranking of trading volume over preceding fiscal year ending in earnings announcement quarter, and (3) an indicator variable equals 1 if price at the end of earnings announcement quarter is greater than $10, and 0 otherwise.

variable and related interactions). In Column (2), we report the results from estimating Equation (6) separately for large traders (i.e., excluding the $EXBMS\_Small_{it}$ variable and related interactions), and in Column (3), we report the results including both the small trade and large trade variables.

In Column (1), the coefficients on *RW* and *AF* are positive and significant, verifying that the drift is associated with both RW and AF earnings surprises. In addition, consistent with H3, the coefficient on *RW·EXBMS_Small* is negative and significant ($\theta_5 = -0.044$, $t = -3.20$, $p < 0.01$). As predicted, this evidence suggests that small traders' announcement period trading intensity, conditioned on RW earnings surprises, reduces the magnitude of the RW drift.[26] In contrast, the coefficient on *AF·EXBMS_Small* is insignificant ($\theta_6 = 0.009$, $t = 0.47$, $p = 0.64$), indicating that small traders' announcement period trading intensity does not reduce the AF drift.[27] Finally, consistent with the drifts increasing with higher transaction costs, we find positive and significant coefficient on *RW·TransCost* (0.076, $t = 2.67$, $p < 0.01$) and positive but insignificant coefficient on *AF·TransCost* (0.042, $t = 1.49$, $p = 0.14$).[28] Consistent with H4, Column (2) reports a negative and significant coefficient on *AF·EXBMS_Large* ($\theta_8 = -0.017$, $t = -2.07$, $p = 0.04$). Thus, as predicted, large traders' announcement period trading intensity in the direction of AF earnings surprises reduces the AF drift. Interestingly, the coefficient on *RW·EXBMS_Large* is positive and marginally significant ($\theta_7 = 0.015$, $t = 1.72$, $p = 0.09$), suggesting that when large traders trade more intensely during the earnings-announcement period in the direction of RW earnings surprises, the RW drift is enhanced.[29] As in Column (1), Column (2) also reports positive coefficients on *RW·TransCost* (0.077, $t = 2.71$, $p < 0.01$) and *AF·TransCost* (0.044, $t = 1.54$, $p = 0.12$).

When we run regression Equation (6) with small and large traders together, we obtain similar results. Specifically, in Column (3) the coefficients on *RW·EXBMS_Small* and *AF·EXBMS_Large* are negative and significant ($\theta_5 = -0.046$, $t = -3.29$, $p < 0.01$; $\theta_8 = -0.018$, $t = -2.16$, $p = 0.03$). In sum, this analysis suggests that the RW and AF drifts, at least in part, are attributable to the under-reaction (i.e., incomplete trading) by different groups of investors with alternative earnings expectations models.

---

[26] Hirshleifer et al. (2008) find that adding the *main* effect of individual investor trades during the announcement period to a regression of post-announcement returns on the RW earnings surprises does not decrease the power of RW to explain post-announcement returns. We are able to reconcile this result. Similar to Hirshleifer et al. (2008), we find that the *main* effect for net small trades does not decrease the power of RW earnings surprises to explain post-announcement returns. In contrast to Hirshleifer et al. (2008), we posit that small trades during the announcement period attenuates the RW portion of the drift (i.e., the association between the RW earnings surprise and post-announcement returns). To capture the statistical significance of this attenuation, we therefore *interact* small trades with the RW earnings surprises (and also include small trades as a main effect) in the regression of post-announcement returns on the RW earnings surprises (see Equation (6)).

[27] If small traders were to arbitrage the AF drift, then we would expect their trading intensity at the earnings announcements should be positively associated with their expected profits of this strategy (i.e., the magnitude of the AF drift), which would result in a positive relation between AF drift and the interaction term of *AF* and *EXBMS_Small*. The insignificant coefficient on the interaction term *AF* and *EXBMS_Small* in Table 4 is inconsistent with arbitrage by small traders.

[28] Evidence that the RW drift increases with transaction cost is consistent with transaction costs providing one explanation for why large traders do not completely eliminate the RW drift.

[29] Ke and Ramalingegowda (2005) and Ali et al. (2007) find evidence consistent with certain institutional investors exploiting the RW drift. A significant positive coefficient on *RW·EXBMS_Large* is consistent with large traders initiating trades around earnings announcements to profit from the RW drift. Specifically, in anticipation of the RW drifts, some sophisticated large traders may trade at the earnings announcements based on the RW surprises. Subsequently, they can trade against the orders of small traders for a profit during the post-earnings-announcement period. The trading intensity of large traders at the earnings announcements should be positively associated with their expected profits of this strategy (i.e., the magnitude of the RW drift), which would result in a positive relation between RW drift and the interaction term of *RW* and *EXBMS_Large*. We note, however, that based on the results in Panel A, Table 3, during the announcement period large traders *on average* do not trade on the RW earnings surprises (the coefficient on *RW* is insignificant). Therefore, as one might expect, arbitrage trading on the RW signal does not seem to be a significant motive for the *average* large trader at earnings announcements.

VI. EVIDENCE OF ADDITIONAL DIFFERENCES BETWEEN THE RW AND AF DRIFTS

We conduct several analyses to further explore and contrast the RW and AF drifts. While our evidence suggests that the RW and AF drifts are attributable to distinct investor groups that under-react to RW and AF earnings surprises, we expect the nature of these two drifts to be qualitatively different because large traders are considered relatively more sophisticated than small traders (e.g., as evidenced by their more sophisticated earnings expectation model). In this section we provide additional analyses that investigate the persistence of trading associated with each drift, the predictability of post-announcement trades based on prior earnings surprises, the roles of analyst forecast heterogeneity and analyst forecast revisions after earnings announcements in explaining the drifts, and how the RW and AF drifts have changed in recent years.

Persistence of Trading after the Earnings Announcements

While we find that small and large investors under-react to distinct forms of earnings surprises, we further investigate whether the persistence of investors' post-announcement trading varies predictably with investor sophistication. To the extent that small traders are relatively less sophisticated than large traders as prior studies argue (Bhattacharya 2001; Bhattacharya et al. 2007; Malmendier and Shanthikumar 2007), we expect small traders' trading on information in RW earnings surprises to continue longer during the post-earnings-announcement period and large traders' trading on information in AF earnings surprises to concentrate at the early stage of the post-announcement period.

To test our expectation, we divide the 60-day post-earnings-announcement period into three 20-day periods. We run the same regression Equation (5) for small and large traders during these three 20-day sub-periods. We expect that large traders' activities are more concentrated in the first 20-day sub-period, while small traders' activities are more evenly distributed among these three sub-periods.

Consistent with expectations, results in Panel A of Table 5 show that small traders' activities do not mainly concentrate in the early sub-period. In particular, the coefficient on *RW* is positive and significant in all three 20-day periods ($\beta_1 = 0.023$, $t = 9.59$, $p < 0.01$ for the first 20-day period, $\beta_1 = 0.033$, $t = 12.80$, $p < 0.01$ for the second 20-day period, and $\beta_1 = 0.034$, $t = 11.70$, $p < 0.01$ for the third 20-day period). In fact, it appears that small traders' trading activities are persistent and even become more intensified over time. Indeed, using a stacked regression including each 20-day period, we find that the coefficients for the second and third 20-day periods are significantly larger than the coefficient for the first 20-day period ($t = 4.36$, $p < 0.01$ and $t = 3.78$, $p < 0.01$, respectively).[30] Panel B reports results for large traders. The coefficients on *AF* are positive and decline in magnitude ($\beta_2 = 0.018$, $t = 3.27$, $p < 0.01$ for the first 20-day period, $\beta_2 = 0.014$, $t = 2.46$, $p = 0.01$ for the second 20-day period, and $\beta_2 = 0.010$, $t = 2.05$, $p = 0.04$

[30] We conjecture that feedback-loop trading (De Long et al. 1990a; 1990b) and/or increased attention (Barber and Odean 2008) cause small traders' trading intensity to increase in subperiods after the earnings announcements. Feedback-loop trading suggests that a subset of small traders trade in an early subperiod and the price changes due to their trading cause another subset of small traders to trade in a later subperiod. Increased attention suggests that a subset of small traders trade more when the stock catches more of their attention as other investors trade more. To determine which conjecture more likely explains the increased small traders' trading intensity after earnings announcements, we regress small traders' trading during the third subperiod [+46, +65] on the two forms of earnings surprises, *RW* and *AF*, the cumulative abnormal return during [+6, +45] $CAR_{+6,+45}$ (or trading volume during [+6, +45] $Vol_{+6,+45}$) and its interaction with the random-walk-based earnings surprise $RW \times CAR_{+6,+45}$ (or $RW \times Vol_{+6,+45}$). The coefficients on $RW \times CAR_{+6,+45}$ and $RW \times Vol_{+6,+45}$ are positive and significant, supporting both the feedback-loop trading conjecture and the increased attention conjecture. However, when $RW \times CAR_{+6,+45}$ and $RW \times Vol_{+6,+45}$ are included simultaneously in the model, the effect of $RW \times CAR_{+6,+45}$ becomes insignificant while the effect of $RW \times Vol_{+6,+45}$ remains positive and significant, suggesting that the attention-based trading effect is likely more dominant.

TABLE 5

Sub-Period Regressions of Excess Net-Buy for Small and Large Trades

Panel A: Small Trades

	[+6, + 25]	[+26, +45]	[+46, +65]
Intercept	0.011	−0.003	0.004
	(7.05)***	(−1.74)**	(2.44)***
RW	0.023	0.033	0.034
	(9.59)***	(12.80)***	(11.70)***
AF	−0.010	−0.007	−0.000
	(−3.91)***	(−2.37)**	(−0.11)
R^2 (%)	0.11	0.20	0.20
Obs.	73,469	73,469	73,469

Panel B: Large Trades

	[+6, +25]	[+26, +45]	[+46, +65]
Intercept	0.005	0.013	0.018
	(1.39)	(4.13)***	(5.09)***
RW	−0.013	−0.020	−0.022
	(−2.47)**	(−4.17)***	(−3.98)***
AF	0.018	0.014	0.010
	(3.27)***	(2.46)**	(2.05)**
R^2 (%)	0.02	0.03	0.03
Obs.	73,469	73,469	73,469

*, **, *** Represent significance levels at less than two-tailed 10 percent, 5 percent, and 1 percent, respectively.
All t-statistics in parentheses are adjusted for clustering by firm (Rogers 1993).

Panel A and Panel B of Table 5 shows regression of average excess net-buy of small trades and average net-buy of large trades during three sub-periods following earnings announcements, [+6, +25], [+26, +45], and [+46, +65]. *RW* is the decile rank of seasonal random-walk earnings surprise converted to [0, 1], and *AF* is the decile rank of analyst-based earnings surprise converted to [0, 1]. Seasonal random walk-earnings surprise is defined as I/B/E/S actual EPS for quarter *t* minus I/B/E/S actual EPS for quarter *t*−4, deflated by price per share at the beginning of quarter *t*. Analyst-based earnings surprise is defined as I/B/E/S actual EPS for quarter *t* minus most recent forecast for quarter *t*, deflated by price per share at the beginning of quarter *t*. Small trade is defined as a trade at maximum $5,000. Large trade is defined as a trade at minimum $30,000.

for the third 20-day period). Using a stacked regression including each 20-day period, we find that the coefficients for the second and third 20-day periods are significantly smaller than the coefficient for the first 20-day period ($t = -1.67$, $p = 0.09$ and $t = -1.83$, $p = 0.07$, respectively). Thus, consistent with our expectation, large traders' trades are more concentrated during the early stage of the post-announcement period.[31]

Trading around Earnings Announcements and Prior Earnings Surprises

If small traders fixate on RW earnings surprises and are unaware of the serial correlations of RW earnings surprises, then we expect that small trades around earnings announcements are

[31] We find that the association between large trades and AF surprises is no longer significant ($p = 0.18$) in the [+46, +65] period when we use a stronger cut on outliers (top and bottom 2 percent). Hence, it appears that the significant association between large trades and AF surprises over the [+46, +65] period is driven by a small number of outliers.

associated with lagged RW earnings surprises (Bernard and Thomas 1990). In contrast, if large traders are more sophisticated and their under-reaction is relatively short-lived, then their trades around earnings announcements are less likely to be associated with lagged AF earnings surprises. To examine the associations of trades at earnings announcements and lagged earnings surprises, we run the following regression:

$$EXBMS\_Small_{it}/EXBMS\_Large_{it} = \delta_0 + \delta_1 RW_{it-q} + \delta_2 AF_{it-q} + \varepsilon_{4it}, \quad q = 1,2,3, \text{ or } 4, \tag{7}$$

where q represents lagged one to four quarters.[32]

Table 6 reports the results from estimating Equation (7). We find that average excess net-buy for small traders around earnings announcements are positively associated with RW_{t-1} (0.043, t = 9.33, p < 0.01), RW_{t-2} (0.027, t = 5.84, p < 0.01), RW_{t-3} (0.024, t = 5.21, p < 0.01), and RW_{t-4} (0.017, t = 3.78, p < 0.01). These results suggest that small trades around one- to four-quarter-ahead earnings announcements are predictable based on prior RW earnings surprises. On the other hand, we find that excess net-buys for large traders are not associated with prior AF earnings surprises, indicating that large trades around future earnings announcements cannot be predicted based on prior AF earnings surprises. This evidence suggests that the RW drift (but not AF drift) is largely explained by small or unsophisticated traders' failure to recognize the time-series properties of earnings, consistent with investor naiveté.

Heterogeneous Forecasts and Post-Announcement Trading

To further differentiate the RW and AF drifts, we investigate whether the AF drift and post-announcement trades by large traders associated with the AF earnings surprise are more pronounced when the AF earnings surprise is more difficult to interpret (i.e., a longer price discovery process when earnings surprises are more difficult to interpret). Prior research finds that the AF drift is greater when analysts' forecasts are more heterogeneous (Liang 2003), suggesting that investors have difficulty interpreting AF earnings surprises when analysts' forecasts are less correlated. Based on Liang (2003), we estimate the following regression model to test the effects of forecast heterogeneity on drift and post-announcement trading:

$$CAR_{it}/EXBMS\_Large_{it}/EXBMS\_Small_{it} = \gamma_0 + \gamma_1 RW_{it} + \gamma_2 AF_{it} + \gamma_3 AF_{it} \cdot Hetero_{it} + \gamma_4 Hetero_{it} + \varepsilon_{6it} \tag{8}$$

where *Hetero* equals 1 minus the BKLS analyst consensus (i.e., Barron et al. 1998). Specifically, we compute *Hetero* as:

$$1 - \frac{SE - D/N}{(SE - D/N) + D} \tag{9}$$

where D is the forecast dispersion, the sample variance of one-year-ahead annual earnings forecasts reported within 30 days following the earnings announcement; N is the number of those forecasts; and SE is the squared difference between the actual annual EPS and the mean of those

[32] We start with a replication of the serial correlations of RW and AF earnings surprises (Bernard and Thomas 1990; Abarbanell and Bernard 1992) for our sample. Consistent with Bernard and Thomas (1990), we find a "+, +, +,−" pattern for first- to fourth-order serial correlations among RW earnings surprises with declining magnitudes (replication results not tabulated). Consistent with Abarbanell and Bernard (1992), we find positive serial correlations among AF earnings surprises for first- to fourth-order serial correlations with declining magnitudes (replication results not tabulated). We also replicate prior results that future returns around earnings announcements are predictable by regressing three-day *CAR* regressions on lagged RW earnings surprises. We find that three-day *CAR* is positively associated with lagged one and two quarters' RW earnings surprises and lagged one to three AF earnings surprises. Further, the coefficients on RW and AF earnings surprises decline in magnitude from lagged one to lagged four quarters (replication results not tabulated).

TABLE 6

Excess Net-Buy for Small and Large Trades and Lagged Earnings Surprises

	Small Trades				Large Trades			
Intercept	0.023	0.029	0.032	0.037	0.032	0.037	0.035	0.036
	(8.27) ***	(10.60) ***	(10.96) ***	(12.83)***	(5.23)	(5.92)	(5.37)	(5.85)
RW_{t-1}	0.043	—	—	—	−0.018	—	—	—
	(9.33) ***				(−1.84) *			
RW_{t-2}	—	0.027	—	—	—	−0.016	—	—
		(5.84) ***				(−1.63)		
RW_{t-3}	—	—	0.024	—	—	—	−0.014	—
			(5.21) ***				(−1.44)	
RW_{t-4}	—	—	—	0.017	—	—	—	−0.010
				(3.78)***				(−1.01)
AF_{t-1}	0.004	—	—	—	0.016	—	—	—
	(1.05)				(1.61)			
AF_{t-2}	—	0.008	—	—	—	0.002	—	—
		(1.28)				(0.16)		
AF_{t-3}	—	—	0.005	—	—	—	0.007	—
			(1.07)				(0.73)	
AF_{t-4}	—	—	—	0.001	—	—	—	−0.001
				(0.34)				(−0.15)
R^2 (%)	0.18	0.08	0.06	0.03	0.01	0.00	0.00	0.00
n	63,223	62,526	61,896	61,665	63,223	62,526	61,896	61,665

*, **, *** Represent significance levels at less than two-tailed 10 percent, 5 percent, and 1 percent, respectively.

All t-statistics in parentheses are adjusted for clustering by firm (Rogers 1993).

Table 6 shows regressions of average excess net-buy of small and large trades during [−1, +1] earnings announcement window on lag earnings surprises. RW_{t-q} (q = 1, 2, 3 or 4) is the decile rank of seasonal random-walk earnings surprise for quarter $t-q$ converted to [0, 1]. AF_{t-q} (q = 1, 2, 3 or 4) is the decile rank of analyst-based earnings surprise for quarter $t-q$ converted to [0, 1]. Seasonal random-walk-earnings surprise is defined as I/B/E/S actual EPS for quarter t minus I/B/E/S actual EPS for quarter t–4, deflated by price per share at the beginning of quarter t. Analyst-based earnings surprise is defined as I/B/E/S actual EPS for quarter t minus most recent forecast for quarter t, deflated by price per share at the beginning of quarter t. Small trade is defined as a trade at maximum \$5,000. Large trade is defined as a trade at minimum \$30,000.

forecasts. If there is only one forecast available, then we set *Hetero* to 0 (i.e., no heterogeneous information contained in analysts' forecasts). Consistent with Liang (2003), we transform *Hetero* into deciles scaled between [0, 1]. We predict that the coefficient for *AF·Hetero* should be positive in both the *CAR* and *EXBMS_Large* regressions if the AF drift and post-announcement trading on AF earnings surprises by large traders is more pronounced when analysts' forecasts are more heterogeneous.

Results in Table 7 are consistent with our expectation. Specifically, we find that the coefficient on *AF·Hetero* is positive and significant in the large traders' trading regression ($\gamma_3 = 0.009$, t = 2.67, p = 0.01), indicating that large traders' post-earnings-announcement period trading associated with AF earnings surprises is more pronounced when analysts' forecasts are more heterogeneous. Similar to Liang (2003), we also find a positive coefficient on *AF·Hetero* in the *CAR*

TABLE 7

Relation between Heterogeneity of Analysts' Forecasts and Returns and Excess Net-Buy during the Post-Earnings-Announcement Period

	60-Day Post-Announcement Period		
	DRIFT	**Small Trades**	**Large Trades**
Intercept	−0.047	0.005	0.005
	(−12.68)***	(3.67)***	(3.67)***
RW	0.041	0.030	−0.025
	(6.62)***	(14.26)***	(−6.96)***
AF	0.012	−0.011	0.006
	(1.51)	(−3.41)***	(1.90)*
AF·Hetero	0.056	−0.030	0.009
	(5.83)***	(−8.02)***	(2.67)***
Hetero	0.016	−0.002	0.001
	(14.94)***	(−4.18)***	(2.02)**
R^2 (%)	0.32	0.39	0.37
Obs.	66,726	66,729	66,726

*, **, *** Represent significance levels at less than two-tailed 10 percent, 5 percent, and 1 percent, respectively.

All t-statistics in parentheses are adjusted for clustering by firm (Rogers 1993).

Table 7 shows regressions of cumulative abnormal return, average excess net-buy of small trades, and excess average net-buy of large trades during [+5, +65] window following earnings announcements. Abnormal return during the post-earnings-announcement period is defined as the firm return in excess of the corresponding Fama-French size and book-to-market six-portfolio benchmark return. *RW* is the decile rank of seasonal random-walk earnings surprise converted to [0, 1], and *AF* is the decile rank of analyst-based earnings surprise converted to [0, 1]. Seasonal random-walk earnings surprise is defined as I/B/E/S actual EPS for quarter *t* minus I/B/E/S actual EPS for quarter *t*–4, deflated by price per share at the beginning of quarter *t*. Analyst-based earnings surprise is defined as I/B/E/S actual EPS for quarter *t* minus most recent forecast for quarter *t*, deflated by price per share at the beginning of quarter *t*. Following Liang (2003), *Hetero* equals 1 minus the BKLS analyst consensus (i.e., Barron et al. 1998). Specifically, we compute *Hetero* as:

$$1 - \frac{SE - D/N}{(SE - D/N) + D}$$

where *D* is the forecast dispersion, the sample variance of one-year-ahead annual earnings forecasts reported within 30 days following earnings announcement; *N* is the number of those forecasts; *SE* is the squared difference between the actual annual EPS and the mean of those forecasts. If there is only one forecast available, *Hetero* is defined as 0 (i.e., no heterogeneous information contained in analysts' forecasts). Consistent with Liang (2003), *Hetero* is transformed into deciles and converted to [0, 1]. Small trade is defined as a trade at maximum $5,000. Large trade is defined as a trade at minimum $30,000.

regression ($\gamma_3 = 0.056$, t = 5.83, $p < 0.01$). This evidence is consistent with investors and analysts requiring more post-announcement information or price discovery to interpret earnings surprises when analyst forecasts are more heterogeneous. Finally, consistent with the pattern in our previous analyses, the results for small traders are opposite to the results for large traders. Specifically, we find a negative and significant coefficient for *AF·Hetero* in the small trader regression ($\gamma_3 = -0.030$, t = −8.02, $p < 0.01$).

Effect of Analyst Forecast Revisions during the Post-Announcement Period

We next investigate how analyst forecast revisions in the post-earnings-announcement period affect large and small traders' post-announcement trading on earnings surprises. We propose that large traders' under-reaction is corrected in the post-announcement period through a price discovery process and offer analyst forecast revisions as one source of information discovery (Gleason and Lee 2003). Specifically, we conjecture that post-announcement trades on AF earnings surprises may be at least partially explained by analyst forecast revisions in the post-announcement period that are correlated with AF earnings surprises. Because our previous results suggest that small traders are less likely to trade on AF earnings surprises, we expect that analyst forecast revisions are less likely to explain small traders' post-announcement trades. We estimate the following regression to determine whether large and small trades on RW and AF earnings surprises in the post-announcement period are partially explained by post-announcement analyst forecast revisions:

$$DRIFT_{it}/EXBMS\_Small_{it}/EXBMS\_Large_{it} = \lambda_0 + \lambda_1 RW_{it} + \lambda_2 AF_{it} + \lambda_3 Rev_{it} + \varepsilon_{5it}, \tag{10}$$

where *Rev* is the decile rank of analyst forecast revisions during the post-earnings-announcement period. We compute an analyst's forecast revision as an analyst's most recent forecast issued during [+46, +65] window minus the *same* analyst's earliest forecast issued during the [+6, +25] window, scaled by lag stock price (Gleason and Lee 2003). To aggregate analyst-level revisions to firm-level, we average individual analysts' revisions for each announcement event. We code forecast revisions as 0 if the I/B/E/S Detail file reports no associated forecasts.

We report regression results for Equation (10) in Panel A of Table 8.[33] In the regression with post-announcement stock returns as the dependent variable, the coefficient on *Rev*, is positive, 0.273, and significant (t = 51.57, $p < 0.01$). As expected, this evidence suggests that market returns incorporate information contained in analyst forecast revisions. Controlling for post-announcement revisions causes the coefficient on *RW* to drop from 0.040 (t = 6.73, $p < 0.01$) to 0.013 (t = 2.30, p = 0.02; the coefficient decrease is significant at $p < 0.01$) and the coefficient on *AF* to drop from 0.043 (t = 7.47, $p < 0.01$) to 0.020 (t = 3.51, $p < 0.01$; the coefficient decrease is significant at $p < 0.01$). This evidence suggests that the information captured by analyst forecast revisions explains substantial portions of the RW and AF drifts.

When we use excess buy-sell order imbalance of large traders as the dependent variable, the coefficient on *Rev* is positive, 0.027, and significant (t = 12.93, $p < 0.01$), consistent with large traders incorporating the information contained in analyst forecast revisions in their trades. With the inclusion of revisions, the coefficient on *AF* declines from 0.015 (t = 4.30, $p < 0.01$) to 0.008 (t = 3.37, $p < 0.01$), a 46.67 percent decrease (significant at $p < 0.01$), whereas the coefficient on *RW* remains largely unchanged. This evidence suggests that much of the post-announcement trading by large traders based on AF earnings surprises is attributable to information contained in

---

[33] The Pearson correlation between *RW* and *Rev* is 0.1234 and between *AF* and *Rev* is 0.1147, consistent with prior research showing an association between post-earnings-announcement analyst forecast revisions in the post-announcement period and both random-walk earnings changes and analyst-based earnings surprises (Abarbanell and Bernard 1992).

TABLE 8

Effects of Analyst Forecast Revisions Following Earnings Announcements on Drifts and Trades

Panel A: 60-Day Post-Announcement Period

	DRIFT		Small Trade		Large Trade	
Intercept	−0.047	−0.113	0.004	0.006	0.012	0.009
	(−13.29)***	(−30.33)***	(3.46)***	(4.68)***	(5.84)***	(5.71)***
RW	0.040	0.013	0.030	0.031	−0.018	−0.020
	(6.73)***	(2.30)**	(15.21)***	(15.43)***	(−5.77)***	(−8.58)***
AF	0.043	0.020	−0.005	−0.005	0.015	0.008
	(7.47)***	(3.51)***	(−2.58)***	(−2.26)**	(4.30)***	(3.37)***
Rev		0.273		−0.008		0.027
		(51.57)***		(−4.02)***		(12.93)***
R^2 (%)	0.26	4.66	0.28	0.30	0.16	0.29
Obs.	73,469	73,469	73,469	73,469	73,469	73,469

Panel B: Large Trades in Three 20-Day Sub-Post-Announcement Periods

	[+6, +25]		[+26, +45]		[+46,+65]	
Intercept	0.005	0.002	0.013	0.006	0.018	0.005
	(1.39)	(0.60)	(4.13)***	(1.96)***	(5.09)***	(1.40)
RW	−0.013	−0.014	−0.020	−0.025	−0.022	−0.026
	(−2.47)**	(−2.63)***	(−4.17)***	(−4.96)***	(−3.98)***	(−4.90)***
AF	0.018	0.017	0.014	0.009	0.010	0.005
	(3.27)***	(3.14)***	(2.46)**	(1.72)*	(2.05)**	(1.46)
Rev		0.018		0.032		0.043
		(4.44)***		(7.50)***		(8.99)***
R^2 (%)	0.02	0.04	0.03	0.10	0.03	0.14
Obs.	73,469	73,469	73,469	73,469	73,469	73,469

*, **, *** Represent significance levels at less than two-tailed 10 percent, 5 percent, and 1 percent, respectively.
All t-statistics in parentheses are adjusted for clustering by firm (Rogers 1993).

(continued on next page)

Panel A of Table 8 shows regression of cumulative abnormal return, average excess net-buy of small trades, and average excess net-buy of large trades during [+6, +65] window following earnings announcements. Abnormal return during the post-earnings-announcement period is defined as the firm return in excess of the corresponding Fama-French size and book-to-market six-portfolio benchmark return. Small trade is defined as a trade at maximum \$5,000. Large trade is defined as a trade at minimum \$30,000. *RW* is the decile rank of seasonal random-walk earnings surprise converted to [0, 1], and *AF* is the decile rank of analyst-based earnings surprise converted to [0, 1]. Seasonal random-walk earnings surprise is defined as I/B/E/S actual EPS for quarter t minus I/B/E/S actual EPS for quarter t–4, deflated by price per share at the beginning of quarter t. Analyst-based earnings surprise is defined as I/B/E/S actual EPS for quarter t minus most recent forecast for quarter t, deflated by price per share at the beginning of quarter t. We compute an analyst's forecast revision as an analyst's most recent forecast issued during [+46, +60] window minus the same analyst's earliest forecast issued during the [+6, +25] window, scaled by lag stock price. To aggregate analyst-level revisions to firm-level, we average individual analysts' revisions for each announcement event. *Rev* is the decile rank (converted to [0, 1]) of this variable. Forecast revisions are coded as 0 if no associated forecasts are reported in the I/B/E/S Detail file.

Panel B of Table 8 shows regression of excess net-buy of large trades during [+6, +25], [+26, +45], and [+46, +65] subperiods. In Panel B, we compute an analyst's forecast revision as an analyst's most recent forecast issued during the respective 20-day subperiod minus the same analyst's previous most recent forecast, scaled by lag stock price. We then aggregate analyst-level revisions to firm-level by averaging individual analyst forecast revisions. *Rev* is the decile rank (converted to [0, 1]) of this variable.

analyst forecast revisions that continue to be correlated with AF earnings surprises well after the earnings announcements. This evidence is also consistent with the AF drift being largely attributable to a longer price discovery process by large traders, with analyst forecast revisions serving as one source of information discovery. In the small-trade regression, the coefficients on *RW* and *AF* are unchanged, while the coefficient on *Rev* is negative and significant. This evidence suggests that post-announcement small trades associated with RW earnings surprises are not explained by information contained in analyst forecasts.[34]

Panel B of Table 8 presents results of large trades in the three 20-day subperiods. In this analysis, we compute an analyst's forecast revision as an analyst's most recent forecast issued during the respective 20-day subperiod minus the same analyst's previous most recent forecast, scaled by lag stock price. We then aggregate analyst-level revisions to firm-level by averaging individual analyst forecast revisions and calculate *Rev* as the decile rank of the firm-level amounts. We find that the association between large trades and AF surprises becomes less significant during the [+26, +45] period and completely disappears during the [+46, +65] period after we control for *Rev*, whereas the association between large trades and *Rev* remains significant across each subperiod. This evidence is consistent with price discovery through analyst forecast revisions explaining the association between AF surprises and large trades well after earnings announcements.[35]

Market and Trading Reactions to Earnings Surprises Over Time

In recent years, analyst forecasts have become more readily available to all investors, which may result in small traders shifting away from RW earnings expectations (Brown and Caylor 2005). Likewise, in recent years, the analyst forecast drift appears to have received more attention, which could result in decreased AF drift as investors either arbitrage the drift or simply react more thoroughly to AF earnings surprises at earnings announcements. We explore these possibilities by introducing a time trend indicator into the following regression model:

$$DRIFT_{it}CAR_{it}/EXBMS\_Small_{it}/EXBMS\_Large_{it} = \varphi_0 + \varphi_1 RW_{it} + \varphi_2 AF_{it} + \varphi_3 RW_{it} \cdot Recent_{it} + \varphi_4 AF_{it} \cdot Recent_{it} + \varphi_5 Recent_{it} + \varepsilon_{7it} \quad (11)$$

where *Recent* is an indicator variable that equals 1 if an earnings announcement occurs after 1999, and 0 otherwise.[36]

Table 9 reports the results from this analysis. During the three-day earnings-announcement period, when we use abnormal return as the dependent variable, the coefficient on *RW·Recent* is insignificant, while the coefficient on *AF·Recent* is positive and significant (0.009, $t = 4.36$, $p < 0.01$), suggesting the market reacts more to AF earnings surprises in more recent years, consistent with Brown and Caylor (2005). When we use excess net-buy of small traders as the dependent variable, the coefficient on *RW·Recent* is negative and significant (-0.066, $t = -8.01$, $p < 0.01$)

---

[34] We obtain similar results in tests in which we include three separate analyst forecast revision variables defined as the difference between the mean value of all annual forecasts issued during the second half of the trading month (e.g., the 1st, 2nd, or 3rd month following the earnings announcement) and the mean value of annual forecasts issued during the first half of the respective month.

[35] To provide further evidence that the AF drift may reflect a price discovery process through analyst forecast revisions after earnings announcements, we sort sample firms into quintile portfolios based on forecast revision frequency and examine how the AF drift varies across these portfolios in the post-announcement period. Consistent with a price discovery story, we find that when analysts' forecast revision frequency increases across Quintiles 2 through 5, the AF drift becomes more pronounced.

[36] We use the year 1999 to split the sample for two reasons. First, this split roughly divides our sample in half. Second, CUSUM statistics (Han and Park 1989) based on recursive least-squares regressions of the drift indicate a structural shift in 1999.

TABLE 9

Changes in the Market and Trade Reactions to Earnings Surprises over Time

	Three-Day Announcement Period			60-Day Post-Announcement Period		
	CAR	Small Trades	Large Trades	*DRIFT*	Small Trades	Large Trades
Intercept	−0.028	0.025	−0.061	−0.056	0.001	0.024
	(−27.92)***	(5.07)***	(−7.08)***	(−10.35)***	(0.36)	(8.38)***
RW	0.011	0.090	0.007	0.029	0.059	−0.036
	(6.85)***	(12.42)***	(0.54)	(3.27)***	(15.63)***	(−7.80)***
AF	0.051	−0.000	0.144	0.083	−0.017	0.019
	(31.08)***	(−0.05)	(10.64)**	(9.15)***	(−4.14)***	(1.78)*
*RW * Recent*	0.002	−0.066	−0.005	0.018	−0.053	0.033
	(0.89)	(−8.01)***	(−0.28)	(1.30)	(−12.85)***	(5.28)***
*AF * Recent*	0.009	0.027	−0.067	−0.073	0.020	−0.011
	(4.36)***	(3.06)***	(−3.63)***	(−6.11)***	(4.33)***	(−1.67)*
Recent	−0.006	−0.025	0.064	0.016	0.007	−0.003
	(−4.35)***	(−4.49)***	(5.57)***	(2.32)**	(2.61)***	(−5.22)***
R^2 (%)	5.82	0.98	0.34	0.35	0.58	0.12
n	73,469	73,469	73,469	73,469	73,469	73,469

*, **, *** Represent significance levels at less than two-tailed 10 percent, 5 percent, and 1 percent, respectively.

All t-statistics in parentheses are adjusted for clustering by firm (Rogers 1993).

Table 9 shows regression of cumulative abnormal return, average excess net-buy of small trades, and average excess net-buy of large trades during [−1, +1] earnings announcement window and during [+6, +65] window following earnings announcements. Abnormal return during post-earnings-announcement period is defined as firm return in excess of corresponding Fama-French size and book-to-market six-portfolio benchmark return. *RW* is the decile rank of seasonal random-walk earnings surprise converted to [0, 1], and *AF* is the decile rank of analyst-based earnings surprise converted to [0, 1]. Seasonal random-walk earnings surprise is defined as I/B/E/S actual EPS for quarter *t* minus I/B/E/S actual EPS for quarter *t*–4, deflated by price per share at the beginning of quarter *t*. Analyst-based earnings surprise is defined as I/B/E/S actual EPS for quarter *t* minus most recent forecast for quarter *t*, deflated by price per share at the beginning of quarter *t*. *Recent* is an indicator variable that equals 1 if an earnings announcement occurs after 1999, and 0 otherwise. Small trade is defined as a trade at maximum $5,000. Large trade is defined as a trade at minimum $30,000.

while the coefficient on *AF·Recent* is positive and significant (0.027, $t = 3.06$, $p < 0.01$), suggesting that small traders decrease their reliance on RW earnings expectations and increase their use of analyst forecasts in more recent years (e.g., as analyst forecasts are more accessible). In comparing the sum of the coefficients for *RW* and *RW·Recent* ($0.090 + -0.066 = 0.024$) to the sum of the coefficients for *AF* and *AF·Recent* ($-0.000 + 0.027 = 0.027$), evidence suggest that in the later part of our sample period small traders trade approximately equally on RW and AF earnings surprises ($F = 0.09$, $p = 0.763$). In contrast, large traders continue to trade only on AF earnings surprises, although the association between large trades and the AF earning surprise declines in the later part of the sample period (i.e., the coefficient for *AF·Recent* is negative and significant (−0.067, $t = -3.63$, $p < 0.01$)).

During the post-earnings-announcement period, the magnitude of the AF drift decreases in more recent years (i.e., the coefficient for *AF·Recent* is negative and significant (−0.073, $t = -6.11$, $p < 0.01$) while the RW drift does not change (i.e., the coefficient for *RW·Recent* is insignificant). This evidence is consistent with a more efficient processing of information contained in the AF earnings surprises in more recent years as the AF drift has received more

attention.[37] Given the relatively new discovery and the little attention that the AF drift had received during the earlier years in our sample, we speculate that sophisticated investors may have not been exploiting the AF drift during this time period, which could explain the larger AF drift in earlier years.

When we use excess net-buy of small traders as the dependent variable, we find that small traders' post-announcement trading intensity related to *RW* decreases in more recent years (the coefficient for *RW·Recent* is -0.053, $t = -12.85$, $p < 0.01$), while that related to *AF* increases in more recent years (coefficient for *AF·Recent* is 0.020, $t = 4.33$, $p < 0.01$). These results are consistent with the trading results around earnings announcements (i.e., decreased trading by small traders on RW earnings surprises).[38] However, we continue to find a significant association between small trades and RW earnings surprises in the post-announcement period in the more recent time period (i.e., the sum of *RW* and *RW·Recent* is 0.006, $F = 11.66$, $p < 0.01$).

When we use excess net-buy of large traders as the dependent variable, we find a decrease in trading on AF earnings surprises in the post-earnings-announcement period (i.e., *AF·Recent* is -0.011, $t = -1.67$, $p = 0.09$). Nonetheless, we continue to find a significant association between large trades and AF earnings surprises in the post-announcement period in the more recent time period (i.e., the sum of *AF* and *AF·Recent* is 0.008, $F = 2.98$, $p = 0.09$).

VII. SUMMARY AND CONCLUSION

Recent research finds two distinct post-earnings-announcement drifts associated with RW and AF earnings surprises. We hypothesize that these two drifts are attributable to the trading activities of small and large traders who under-react to different forms of earnings innovations. Consistent with our hypothesis, we find that small (large) traders continue to trade in the direction of RW (AF) earnings surprises after earnings announcements. We corroborate these findings with evidence that when small (large) traders react more thoroughly to RW- (AF-) based earnings surprises during the announcement period, the respective drift attenuates.

In additional tests, we predict and find that the timing of post-announcement trading varies with trade size—i.e., large (and *relatively* more sophisticated) traders end their post-announcement trading more quickly than small (and *relatively* less sophisticated) traders, and that small trades (but not large trades) at earnings announcements are predictable based on prior earnings surprises. Further, we find that large traders' post-announcement trading is more pronounced when AF earnings surprises are more difficult to interpret (i.e., when analysts' forecasts are more heterogeneous) and that large traders' post-announcement trades based on AF earnings surprises are largely explained by the information contained in post-announcement analyst forecast revisions that are generally in the same directions of the earnings surprises. Finally, we find that the AF drift and related large trades during the post-announcement period have decreased in more recent years as the AF drift may have received more attention.

This study makes the following contributions. First, we predict and find that small traders systematically trade in the direction of RW earnings surprises after earnings announcements, whereas large traders' trading during post-earnings-announcement period is in the direction of AF earnings surprises. This evidence suggests that the RW and AF drifts are attributable, at least in

---

[37] The sum of *AF* and *AF·Recent* is 0.010 ($F = 3.82$, $p = 0.08$), suggesting a marginally significant AF drift in more recent years (2000–2005). Though we lack TAQ data for 2006 and 2007, we examined the AF drift for these years and find no evidence of an AF drift, consistent with a declining AF drift in recent years.

[38] To determine if the small trade results are robust to a period when sophisticated traders less frequently split their trades into smaller trades, we re-estimated Equation (11) focusing on the pre-1999 subsample period and re-defining *Recent* equal to 1 if the sample year is between 1996 and 1998, and 0 if the sample year is between 1993 and 1995. Results are similar—e.g., we find positive coefficients for *AF·Recent* in the small trade regressions during the earnings-announcement and the post-earnings-announcement periods.

part, to distinct sets of investors who use different earnings expectation models. Second, we provide evidence that trading activities during the earnings-announcement period affect the post-earnings-announcement drift. To our knowledge, this is the first study to link either *announcement period trading* or *post-announcement period trading* by distinct groups of investors to the RW and AF *drift*. Third, we find that the two drifts are qualitatively different—delayed small trades associated with random-walk-based surprises are consistent with small traders' failure to understand time-series properties of earnings, whereas delayed large trades associated with analyst-based surprises are more consistent with a longer price discovery process. Finally, we find that the AF drift and related large trades during the post-announcement period have decreased in more recent years as the AF drift appears to have received more attention.

REFERENCES

Abarbanell, J., and V. Bernard. 1992. Tests of analysts' over-reaction/under-reaction to earnings information as an explanation for anomalous stock price behavior. *The Journal of Finance* 47 (3): 1181–1207.

———, and B. Bushee. 1997. Fundamental analysis, future earnings, and stock prices. *Journal of Accounting Research* 35 (1): 1–24.

Admati, A. 1985. A noisy rational expectations equilibrium for multi-asset securities markets. *Econometrica* 53 (3): 629–658.

Ali, A., J. Chen, T. Yao, and T. Yu. 2007. Profiting from the post-earnings-announcement drift: Mutual fund trades, market frictions, and market efficiency. Working paper, The University of Texas at Dallas, Kansas State University, University of Rhode Island, and The University of Iowa.

———, A. Klein, and J. Rosenfield. 1992. Analysts' use of information about permanent and transitory earnings components in forecasting annual EPS. *The Accounting Review* 67 (1): 183–198.

Ayers, B., O. Li, and J. Robinson. 2008. Tax-induced trading around the Taxpayer Relief Act of 1997. *The Journal of the American Taxation Association* 30 (1): 77–100.

Ball, R. 1992. The earnings-price anomaly. *Journal of Accounting and Economics* 15 (2–3): 319–345.

Bamber, L. S. 1987. Unexpected earnings, firm size, and trading volume around quarterly earnings announcements. *The Accounting Review* 62 (3): 510–532.

Barber, B., and T. Odean. 2007. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies* 21 (2): 785–818.

Barclay, M., and J. Warner. 1993. Stealth trading and volatility. *Journal of Financial Economics* 34 (3): 281–305.

Barron, O., O. Kim, D. Lim, and D. Stevens. 1998. Using analyst' forecasts to measure properties of analysts' information environment. *The Accounting Review* 73 (4): 421–433.

Bartov, E., S. Radhakrishnan, and I. Krinsky. 2000. Investor sophistication and patterns in stock returns after earnings announcements. *The Accounting Review* 75 (1): 43–63.

Battalio, R., and R. Mendenhall. 2005. Earnings expectations, investor trade size, and anomalous returns around earnings announcements. *Journal of Financial Economics* 77 (2): 289–319.

Bernard, V., and J. Thomas. 1989. Post-earnings-announcement drift: Delayed price response or risk premium? *Journal of Accounting Research* 27 (supplement): 1–36.

———, and ———. 1990. Evidence that stock prices do not fully reflect the implications of current earnings for future earnings. *Journal of Accounting and Economics* 13 (4): 305–340.

Bhattacharya, N. 2001. Investors' trade size and trading responses around earnings announcements: An empirical investigation. *The Accounting Review* 76 (2): 221–245.

———, E. Black, T. Christensen, and R. Mergenthaler. 2007. Who trades on pro forma earnings information? *The Accounting Review* 82 (3): 581–619.

Bhushan, R. 1994. An informational efficiency perspective on the post-earnings-announcement drift. *Journal of Accounting and Economics* 18 (1): 45–65.

Bradshaw, M., B. Bushee, and G. Miller. 2004. Accounting choices, home bias, and U.S. investment in non-U.S. firms. *Journal of Accounting Research* 42 (5): 795–841.

———, M. Drake, J. Myers, and L. Myers. 2010. A re-examination of analysts' superiority over time-series forecasts. Working paper, Boston College, The Ohio State University, and University of Arkansas.

Brown, L., and M. Caylor. 2005. A temporal analysis of quarterly earnings thresholds: Propensities and valuation consequences. *The Accounting Review* 80 (2): 423–440.

———, and K. Kim. 1991. Timely aggregate analyst forecasts as better proxies for market earnings expectations. *Journal of Accounting Research* 29 (2): 382–385.

Bushee, B. 2001. Do institutional investors prefer near-term earnings over long-run value? *Contemporary Accounting Research* 18 (2): 207–246.

Calegari, M., and N. Fargher. 1997. Evidence that prices do not fully reflect the implications of current earnings for future earnings: An experimental markets approach. *Contemporary Accounting Research* 14 (3): 397–433.

Callen, J., O. K. Hope, and D. Segal. 2005. Domestic and foreign earnings, stock return variability, and the impact of investor sophistication. *Journal of Accounting Research* 43 (3): 377–412.

Campbell, J. T. Ramadorai, and T. Vuolteenaho. 2005. Caught on tape: Institutional order flow and stock returns. Working paper, Harvard University, University of Oxford, and Arrowstreet Capital, LP.

Chakravarty, S. 2001. Stealth-trading: Which trader's trades move stock prices? *Journal of Financial Economics* 61 (2): 289–307.

De Franco, G., H. Lu, and F. Vasvari. 2007. Wealth transfer effects of analysts' misleading behavior. *Journal of Accounting Research* 45 (1): 71–110.

De Long, J., A. Shleifer, L. Summers, and R. Waldmann. 1990a. Positive feedback investment strategies and destabilizing rational speculation. *The Journal of Finance* 45 (2): 379–395.

———, ———, ———, and ———. 1990b. Noise trader risk in financial markets. *The Journal of Political Economy* 98 (4): 703–738.

Doyle, J., R. Lundholm, and M. Soliman. 2006. The extreme future stock returns following I/B/E/S earnings surprises. *Journal of Accounting Research* 44 (5): 849–887.

Easley, D., and M. O'Hara. 1987. Price, trade size, and information in securities markets. *Journal of Financial Economics* 19 (1): 69–90.

Gleason, C., and C. Lee. 2003. Analyst forecast revisions and market price discovery. *The Accounting Review* 78 (1): 193–225.

Grossman, S. 1976. On the efficiency of competitive stock markets where traders have diverse information. *The Journal of Finance* 31 (2): 573–585.

———, and J. Stiglitz. 1980. On the impossibility of informationally efficient markets. *The American Economic Review* 70 (3): 393–408.

Han, A. K., and D. Park. 1989. Testing for structural change in panel data: Application to a study of U.S. foreign trade in manufacturing goods. *The Review of Economics and Statistics* 71 (1): 135–142.

Hirshleifer, D., J. Myers, L. Myers, and S. Teoh. 2008. Do individual investors cause post-earnings-announcement drift? Direct evidence from personal trades. *The Accounting Review* 83 (6): 1521–1550.

Ke, B., and K. Petroni. 2004. How informed are actively trading institutional investors? Evidence from their trading behavior before a break in a string of consecutive earnings increases. *Journal of Accounting Research* 42 (5): 895–927.

———, and S. Ramalingegowda. 2005. Do institutional investors exploit the post-earnings-announcement drift? *Journal of Accounting and Economics* 39 (1): 25–53.

Kandel, E., and N. Pearson. 1995. Differential interpretation of public signals and trade in speculative markets. *The Journal of Political Economy* 103 (4): 831–872.

Kimbrough, M. 2005. The effect of conference calls on analyst and market under-reaction to earnings announcements. *The Accounting Review* 80 (1): 189–219.

Kothari, S. P. 2001. Capital markets research in accounting. *Journal of Accounting and Economics* 31 (1–3): 105–231.

Lee, C. 1992. Earnings news and small traders. *Journal of Accounting and Economics* 15 (2–3): 265–302.

———, and B. Radhakrishna. 2000. Inferring investor behavior: Evidence from TORQ data. *Journal of Financial Markets* 3 (2): 83–111.

———, and M. Ready. 1991. Inferring trade direction from intraday data. *The Journal of Finance* 46 (2): 733–746.

Liang, L. 2003. Post-earnings announcement drift and market participants' information processing biases. *Review of Accounting Studies* 8 (2–3): 321–345.

Livnat, J., and R. Mendenhall. 2006. Comparing the post-earnings-announcement drift for surprises calculated from analyst and time series forecasts. *Journal of Accounting Research* 44 (1): 177–205.

Maines, L., and J. Hand. 1996. Individuals' perceptions and misperceptions of time series properties of quarterly earnings. *The Accounting Review* 71 (3): 317–336.

Malmendier, U., and D. Shanthikumar. 2007. Are small investors naive about incentives? *Journal of Financial Economics* 85 (2): 457–489.

Mendenhall, R. 1991. Evidence on the possible underweighting of earnings-related information. *Journal of Accounting Research* 29 (1): 170–179.

Mikhail, M. B., B. R. Walther, and R. H. Willis. 2007. When security analysts talk, who listens? *The Accounting Review* 82 (5): 1227–1253.

Morse, D. 1981. Price and trading volume reaction surrounding earnings announcements: A closer examination. *Journal of Accounting Research* 19 (2): 374–383.

Rogers, W. 1993. *Regression Standard Errors in Clustered Samples*. College Station, TX: StataCorp LP.

Shanthikumar, D. 2004. Small and large trades around earnings announcements: Does trading behavior explain post-earnings-announcement drift? Working paper, Harvard University.

Shleifer, A., and R. Vishny. 1997. The limits of arbitrage. *The Journal of Finance* 52 (1): 35–55.

Stevens, D., and A. Williams. 2004. Inefficiency in earnings forecasts: Experimental evidence of reactions to positive vs. negative information. *Experimental Economics* 7 (1): 75–92.

Walther, B. 1997. Investor sophistication and market earnings expectations. *Journal of Accounting Research* 35 (2): 157–179.

THE ACCOUNTING REVIEW
Vol. 86, No. 2
2011
pp. 417–449

American Accounting Association
DOI: 10.2308/accr.00000019

Discretionary Disclosure in Financial Reporting: An Examination Comparing Internal Firm Data to Externally Reported Segment Data

Daniel A. Bens
The University of Arizona

Philip G. Berger
The University of Chicago

Steven J. Monahan
INSEAD

ABSTRACT: We use confidential, U.S. Census Bureau, plant-level data to investigate aggregation in external reporting. We compare firms' plant-level data to their published segment reports by grouping a firm's plants that share the same four-digit SIC code into a "pseudo-segment." We then determine whether each pseudo-segment is disclosed as an external segment, or whether it is subsumed into a different business unit for external reporting purposes. We show that a pseudo-segment is more likely to be aggregated when the agency and proprietary costs of separately reporting the pseudo-segment are higher and when firm and pseudo-segment characteristics allow for more

We gratefully acknowledge the support for this research that was provided by the National Science Foundation (NSF awards no. SES-0004335 and ITR-0427889). Professor Bens gratefully acknowledges financial support provided by the Frank and Susan Parise Fellowship at The University of Arizona. Professor Berger gratefully acknowledges financial support provided by The University of Chicago Booth School of Business and its Neubauer Family Faculty Fellows program. Professor Monahan gratefully acknowledges financial support provided by the INSEAD Alumni Fund. We have benefited from the comments of two anonymous reviewers, Rebecca Hann, Ole-Kristian Hope, Mary Stanford, and Wayne Thomas (the editor), as well as workshop participants at the 2008 Tel Aviv Conference in Accounting, the U.S. Census Bureau Center for Economic Studies, the 2007 UC-Davis Conference on Financial Markets Research, and workshops at the following universities: The University of Arizona, University of British Columbia, University of Colorado, University of Connecticut, INSEAD, University of Maryland, University of Michigan, The Pennsylvania State University, Texas Christian University, University of Toronto, University of Virginia (Darden School of Business), University of Washington in St. Louis, and University of Wisconsin.

This research was conducted while the authors were Special Sworn Status researchers of the U.S. Census Bureau at the Chicago Census Research Data Center (RDC). Lynn Riggs and Frank Limehouse of the Chicago RDC provided excellent advice and guidance about the project. Research results and conclusions expressed are those of the authors and do not necessarily reflect the views of the Census Bureau. This paper has been screened to ensure that no confidential data are revealed.

Editor's note: Accepted by Wayne Thomas.

Submitted: June 2008
Accepted: July 2010
Published Online: March 2011

discretion in the application of segment reporting rules. For firms reporting multiple external segments, aggregation of pseudo-segments is driven by both agency and proprietary costs. For firms reporting a single external segment, we find no evidence of an agency cost motive for aggregation.

Keywords: *manufacturing plants; micro-level data; segment reporting; discretionary disclosure; agency costs; proprietary costs.*

Data Availability: *All data are available from public sources, except for the Census micro-level data.*

I. INTRODUCTION

We use a Census Bureau database of *confidential*, plant-level data to investigate two questions: What factors motivate managers of publicly traded firms to aggregate segment information when making disclosure decisions? How do private competitors affect both the disclosure decisions and industry-level statistics used to construct explanatory variables?

The Census data give us three advantages over prior research. First, we better observe management's private information endowment. Past segment-disclosure research (e.g., Harris 1998; Botosan and Stanford 2005; Berger and Hann 2007) has examined managers' aggregation decisions by evaluating the amount of segment information in *publicly* disseminated financial statements. A limitation of this approach is that the researcher does not consider the underlying source data that management observe. Because the Census data contain confidential information about a firm's operations that are less aggregated than the information in public filings, we are able to: (1) observe internal activities conducted by the firm, even if these activities are not separately disclosed in the segment footnotes, and (2) more accurately measure variables.

Observing management's private information endowment is made more valuable by the linkage between the agency cost and proprietary cost hypotheses for nondisclosure that we investigate. Agency costs of segment disclosure may arise when segment data provide information about value-reducing aspects of a firm's diversification strategy. If agency costs were the only plausible motive for nondisclosure, then there would be no point in withholding information, as rational market participants would infer that nondisclosure signals the worst possible outcome (Grossman 1981; Milgrom 1981). Thus, a plausible proprietary cost motive is necessary for the agency cost motive to potentially exist. Moreover, the agency and proprietary cost motives for aggregation are reasonable only if outsiders cannot use *publicly available* disclosures to fully unravel these motives. Thus: (1) the agency and proprietary cost motives should be studied together, and (2) data from *publicly disseminated* financial statements have limitations for drawing inferences about either of these motives for nondisclosure. Hence, the confidential Census data at the disaggregated operating level are a key advantage for this study.

The second advantage of the Census data is the ability to examine public and private U.S. establishments, allowing us to more accurately measure industry-level phenomena that affect disclosure decisions. In addition, we are able to develop a new proprietary cost measure that reflects the portion of total industry sales made by private firms. We use this variable to test whether public firms in industries with high concentrations of private competitors are more likely to mimic the nondisclosure policies of their private peers.

We evaluate proprietary and agency cost motives for nondisclosure. Prior studies of segment reporting (e.g., Harris 1998; Botosan and Stanford 2005) have focused on the proprietary cost motive, which posits that nondisclosure occurs to conceal proprietary information from competitors, suppliers, or regulators. Although the proprietary cost motive has received considerable attention, there is no clear consensus regarding its descriptive validity due to three main issues.

First, the evidence is conflicting. For example, consider the evidence regarding the role of industry concentration, a commonly used measure of product market competition. Bamber and Cheon (1998) find that whether a firm *provides* earnings forecasts is negatively related with industry concentration; however, Verrecchia and Weber (2006) find that whether a firm asks the Securities and Exchange Commission (SEC) to *withhold* proprietary information from its filings is also negatively related to industry concentration. Hence, the former study concludes that there is *less* informative disclosure in more concentrated industries, while the latter study concludes that there is *more*.

Second, there are concerns about the way proprietary cost proxies have been measured. Again, consider product market competition proxies. Typically, these proxies are calculated using Compustat data. However, Ali et al. (2009) show that industry concentration measures calculated using Compustat data are weak proxies for total (public and private firm) industry concentration. They show that using Census industry concentration measures often reverses prior results (including research on segment reporting) that were based on Compustat data.

Finally, Berger and Hann (2007) argue that much of the extant evidence consistent with the proprietary cost hypothesis is also consistent with an alternative agency cost hypothesis that disclosures are withheld due to manager-shareholder conflicts of interest. Berger and Hann's (2007) findings are consistent with the agency cost hypothesis, but are at best mixed with respect to the proprietary cost hypothesis.

We develop measures for both proprietary and agency costs and examine the relative strength of each after controlling for several "non-strategic" determinants of disclosure. We use a sample of 1,625 firm-years from 1987, 1992, and 1997, with coverage in both Compustat and the Census's Longitudinal Research Database (LRD). Our analyses are conducted using logistic regressions at the "pseudo-segment" level. We aggregate all of the firms' LRD plants within the same four-digit Standard Industrial Classification (SIC) code together into one pseudo-segment. The dependent variable is an indicator equal to 1 for pseudo-segments with four-digit SIC codes that match either the primary or secondary SIC code of a disclosed segment for that firm on Compustat ("disclosed" pseudo-segments), and 0 otherwise ("hidden" pseudo-segments).

We analyze both single- and multi-segment firms, with our focus on the multi-segment firms for three reasons. First, consistency with prior work. Second, we argue that multi-segment firms have greater motive and opportunity to aggregate. Finally, our empirical results are consistent with the previous argument, as they suggest that managers of multi-segment firms behave more strategically in their aggregation decisions than do managers of single-segment firms.

Our most robust results within the multi-segment sample are that the likelihood a pseudo-segment is disclosed separately is negatively related to: (1) whether the pseudo-segment receives inefficient transfers of funds from the remainder of the firm, and (2) the speed of abnormal profit adjustment exhibited by firms in the pseudo-segment's industry.

The negative relation between inefficient transfers and the likelihood that a pseudo-segment is separately disclosed is consistent with the agency cost motive: managers suppress information about inefficient internal capital transfers. The interpretation of the relation between the likelihood of disclosure and the industry speed of abnormal profit adjustment is less clear-cut. To understand better the relation between the likelihood of disclosure and the industry speed of profit adjustment, we separate our multi-segment sample into observations that have value-reducing diversification programs (i.e., negative excess value firms) and observations that have value-enhancing diversification programs (i.e., nonnegative excess value firms), estimating our model separately for each

of these subsamples.[1] We show that the industry speed of abnormal profit adjustment is negatively related to the probability of pseudo-segment disclosure in the subsample with value-enhancing diversification but not in the subsample with value-reducing diversification. Given that the value-enhancing subsample is less likely to contain firms with agency cost motives for pseudo-segment aggregation, the results support the proprietary cost interpretation discussed in Harris (1998).

Within the multi-segment sample, we also find that pseudo-segment disclosure is positively associated with the pseudo-segment's industry-adjusted profitability and negatively associated with a labor power proxy that equals the ratio of aggregate industry wages to aggregate industry sales. The first result is consistent with the agency cost motive for nondisclosure, as it implies that information about less profitable operations is hidden. The second result is consistent with the proprietary cost motive, as it implies that firms act in their shareholders' interest by withholding information from other rent-seeking stakeholders. While both results are manifest in our main set of tests, they are not robust to some of our sensitivity tests. In contrast, the results related to the industry speed of abnormal profit adjustment and to inefficient transfers are robust.

We also analyze single-segment firms and find results suggesting their managers behave less strategically than those of multi-segment firms when making disclosure decisions. When single-segment firm managers are strategic, their nondisclosure is to avoid revealing information to private competitors. In particular, for our single-segment sample, only one of our agency/proprietary cost proxies is associated with the likelihood of disclosure. This proxy equals the proportion of industry sales attributable to privately held firms, and it is negatively associated with the probability of disclosure. This result offers a potentially new insight about the nature of proprietary costs, as it suggests that publicly traded firms seek to minimize disclosure levels when they compete with privately held firms that have no legal duty to publicly disclose.

Section II discusses our hypotheses. Section III presents our sample selection, data description, and research design. Results are discussed in Sections IV and V, with concluding comments in Section VI.

II. HYPOTHESIS DEVELOPMENT

Based on models such as those in Verrecchia (1983) and Hayes and Lundholm (1996), prior empirical studies have explored whether proprietary information costs reduce disclosure of industry segments via aggregation. The evidence is mixed. Harris (1998) finds that operations in more concentrated industries are less likely to be reported as separate industry segments, but Ali et al. (2009) show that Harris's finding is attributable to the use of Compustat data, which generally exclude privately held firms, to measure concentration. Botosan and Harris (2000) find no association between proprietary costs and voluntary increases in segment disclosure frequency. Botosan and Stanford (2005) find that managers hide profitable segments operating in less competitive industries, but Berger and Hann (2007) generally do not find such results.

Managers may also face agency costs of segment disclosure, because disaggregated segment data can provide information indicative of unresolved agency problems. Segment data provide information about a company's diversification strategy and its transfers of resources across divisions. Prior research finds evidence consistent with internal capital markets in conglomerates transferring funds across segments in a suboptimal manner (Berger and Ofek 1995; Lamont 1997; Shin and Stulz 1998; Rajan et al. 2000). Several studies indicate that diversified firms trade at a

[1] We consider a multi-segment firm's diversification program to be value-reducing (value-enhancing) if the excess value measure developed in Berger and Ofek (1995) is less than zero (nonnegative). The excess value measure represents an estimate of diversification's effect on firm value.

discount relative to stand-alone firms (Lang and Stulz 1994; Berger and Ofek 1995) and that the diversification discount is associated with measures of agency problems (Denis et al. 1997; Berger and Ofek 1999).

Berger and Hann (2003) find that firms moving to more disaggregated segment reporting under SFAS No. 131 experienced an increased diversification discount, implying that managers concealed information about agency problems under SFAS No. 14.[2] The same implication arises from Bens and Monahan (2004), who find that greater voluntary segment disaggregation is associated with a smaller diversification discount. Berger and Hann (2007) test this implication and find evidence consistent with the withholding of segment data being motivated by the desire to conceal agency problems.

III. SAMPLE SELECTION, DATA, AND METHODOLOGY

Sample Selection and Data

Our sample selection begins with all annual firm observations from the years 1987, 1992, and 1997 for which the firm is listed on Compustat's Annual Industrial, Research, or Full Coverage files and also covered within the Longitudinal Research Database (LRD), maintained by the Center for Economic Studies at the Bureau of the Census. The LRD is made up of two databases. The first is the *Census of Manufacturers* (*CM*), which is conducted every five years. The unit of analysis is a manufacturing plant, which the Census refers to as an *establishment*. Extremely small establishments are excluded from the *CM*, although the Census imputes estimates for them based on Internal Revenue Service and Social Security Administration data; all other establishments are required by law to respond truthfully to the *CM* (U S. Code Title 13, §224). The second LRD database is the *Annual Survey of Manufacturers* (*ASM*), which is conducted in all non-*CM* years. Through the end of our sample period, all establishments with more than 250 employees as of the most recent *CM* are included in the *ASM* panel.[3]

We use only the *CM* years to ensure coverage of nearly all of establishments in each industry. Because many of our independent variables are industry-based, this decreases measurement error. It also makes our research more comparable to studies that use *publicly available* industry measures from the *CM* available on the Census website. Although such measures are limited to a few industry-level variables such as the four-firm concentration index and the Herfindahl concentration index, recent studies such as Ali et al. (2009) and Tang (2009) demonstrate the importance of including private firms in these measures.

The sample begins in 1987 because that is the first *CM* year for which we have Compustat segment data and for which Compustat reports segment SIC codes. The sample ends in 1997 because that is the last year in which firms report segments under SFAS No. 14 rules and also the last year in which Census classified industries using SIC codes (NAICS codes are now used). Ending the sample with the completion of the SFAS No. 14 regime allows us to restrict our attention to *industry-based*, as opposed to *internal-management*, segment reporting. Nevertheless, our findings are likely to also apply to the SFAS No. 131 regime, as it continues to allow managers to have discretion in aggregating plant-level data into reportable segments and because the vast majority of the internal-management segments under SFAS No. 131 turn out to be based on either industry or geography. More importantly, our goal is to use segment reporting to examine discre-

---

2 SFAS No. 14 is FASB Statement No. 14, *Financial Reporting for Segment of a Business Enterprise* (FASB 1976). SFAS No. 131 is FASB Statement No. 131, *Disclosures about Segments of an Enterprise and Related Information* (FASB 1997).

3 Comprehensive descriptions of the LRD are provided by McGuckin and Pascoe (1988) and the website of the Center for Economic Studies (http://www.ces.census.gov).

tion in financial reporting aggregation generally. The greater discretion available to managers under SFAS No. 14 is thus a desirable feature for our research.

We aggregate all of the firm's establishments that operate within the same four-digit SIC code together into one "pseudo-segment." Sample firms must have *at least two* pseudo-segments (i.e., operate multiple plants with at least two unique four-digit SIC codes). This ensures that all sample firms have some potential for aggregation of pseudo-segments into financial reporting segments, as well as the opportunity to transfer resources across pseudo-segments.

Because the LRD covers only U.S. manufacturing establishments, we require that the firm be domiciled in the U.S. and that its primary SIC code on Compustat fall within the manufacturing sector (i.e., between 2000 and 3999). Even so, many firms have a portion of their production that occurs either outside the U.S. or outside of manufacturing. Moreover, while Compustat correctly eliminates intersegment sales within a firm, some double-counting of interplant sales occurs in the Census data. Therefore, the last step in our sample selection process eliminates observations for which there is a poor match between the firm's sales per the LRD and its sales per Compustat. In particular, we remove observations when the ratio of the firm's LRD *total value of shipments* (*TVS*) to Compustat sales is less than 0.75 or greater than 1.25. *TVS* is the Census term for sales.[4] The ratio of *TVS* to sales is below 0.75 if the firm has significant non-manufacturing sales or it has production facilities outside the U.S. generating sales. The ratio exceeds 1.25 if intersegment sales have not been eliminated properly by Census.

Dependent Variable

Following Harris (1998), we evaluate a dichotomous dependent variable, *MATCH*, that captures whether a pseudo-segment is disclosed as a separate external segment. In particular, *MATCH* equals 1 if the pseudo-segment's four-digit SIC code per the Census matches the primary or secondary SIC code of one of the firm's line-of-business segments per Compustat, and 0 otherwise. Under SFAS No. 14, enterprises were required to classify line-of-business segment information using the *industry approach*. Thus, we assess the extent to which a four-digit SIC-based classification of industries for all the firm's plant-level data match the externally reported industry segments chosen by management.[5]

Segment disclosure has been a contentious issue for decades. Researchers' interest derives from the fact that, while SFAS No. 14 required disclosure about firms' operations in different industries (paragraph 1), it explicitly stated that no single industry classification system (such as the SIC code) "is, by itself, suitable to determine industry segments for purposes of this Statement" (paragraph 12). Paragraph 12 continues with the following:

> [N]o single set of characteristics is universally applicable in determining the industry segments of all enterprises, nor is any single characteristic determinative in all cases. Consequently, determination of an enterprise's industry segments must depend to a considerable extent on the management of the enterprise.

The discretion afforded to management is what makes segment reporting an interesting disclosure choice. However, this discretion also makes it difficult for researchers to determine which industries *should* be disclosed. In our setting, we rely on further guidance from SFAS No. 14 to

---

[4] *TVS* is defined as the "selling value f.o.b. plant after discounts and allowances and excluding freight charges" (*Value of Product Shipments: 2005—Annual Survey of Manufactures*, Appendix A, issued November 2006).

[5] We are interested only in the line-of-business (LOB) segment reporting choice for two reasons. First, as a practical matter, the LRD data cover only U.S. manufacturing plants and the extent of geographical diversification in these data is thus limited to within-U.S. variation. Second, with the exception of Hope and Thomas (2008), who focus on geographic diversification, the extant literature has focused on agency problems within the context of line-of-business diversification.

generate predictions of unbiased disclosure. Specifically, paragraph 13 states:

> An enterprise's existing profit centers—the smallest units of activity for which revenue and expense information is accumulated for internal planning purposes—represent a logical starting point for determining the enterprise's industry segments.

The definition of *profit center* coincides with Census's choice to collect plant-level data. This, in turn, justifies our maintained assumption that a pseudo-segment (i.e., a group of profit centers sharing the same four-digit SIC code) is the starting point from which management uses its discretion under SFAS No. 14 when deciding whether to aggregate the pseudo-segment with others or to report it separately in the segment footnotes. We are not asserting that each pseudo-segment *should* be reported separately; rather, our maintained assumption is that each pseudo-segment *has the potential* to be reported separately.

Given the FASB explicitly states that four-digit SIC codes are not a sufficient basis for *all* firms to determine their reportable segments, our aggregation of plants by common four-digit SIC code into pseudo-segments is *ad hoc*. On the other hand, the FASB provides no preferable alternative. Moreover, defining industries with SIC codes has a long history in the literature, with Bhojraj et al. (2003) noting that SIC codes have been available since 1939 and that more than 90 percent of studies that use a general industry classification scheme use SIC codes despite their shortcomings. Thus, we believe it reasonable to rely on this classification scheme while also recognizing that the use of four-digit SIC codes may introduce measurement error. Hence, in sensitivity analyses we: (1) report a validity check of our disclosure measure, *MATCH*, using a comparison based on the data set constructed by Berger and Hann (2003; 2007) and (2) conduct a sensitivity check in which we define industries on the basis of three-digit SIC codes.

Model of Managers' Segment Reporting Choice

We examine managers' segment reporting decisions by estimating, at the pseudo-segment level, the logit regression tabulated in Table 4. The dependent variable, *MATCH*, is a dichotomous variable with the value of 1 if the pseudo-segment's four-digit SIC code per the Census matches a primary or secondary SIC code of a Compustat line-of-business segment reported by the firm, and 0 otherwise.

Treatment Variables

I_PROFIT is a ratio that measures abnormal profitability. The numerator equals the pseudo-segment's gross margin percentage less the average gross margin percentage for all establishments in its industry.[6] The denominator is the standard deviation of the gross margin percentage for all establishments in the pseudo-segment's industry. *I_PROFIT* does not fully capture *total* pseudo-segment profitability (i.e., it ignores some expenses and, more importantly, asset turnover). However, by industry-adjusting, we reduce the likelihood of our profit measure having a low correlation with a total profitability measure such as return on assets (ROA). The rationale is that the negative correlation typically observed between total asset turnover and gross margin is smaller within industries than across industries. We do not attempt to calculate an ROA-based measure because the Census data do not contain sufficient information about establishment-level assets.

I_PROFIT is a direct measure of pseudo-segment profits that captures how well a pseudo-segment performs relative to its industry. It is pseudo-segment profits rather than publicly available firm-level (or external segment-level) profits that managers may try to aggregate. Thus, because it is based on confidential information that is not observable by outsiders, *I_PROFIT* is a

---

6 Please refer to the notes to Table 2 for additional details regarding the construction of our treatment and control variables. Unless otherwise noted, the term "industry" refers to a group of firms, establishments, segments, or pseudo-segments that have the same four-digit SIC code.

superior proxy to publicly observable measures of profit. The coefficient relating *I_PROFIT* to disclosure will be positive if the agency cost hypothesis dominates and only strongly performing units tend to be disclosed. Conversely, the coefficient will be negative if the proprietary cost hypothesis dominates and strong performers are hidden.

PROFITADJ is an industry abnormal profit adjustment measure similar to a construct used by Harris (1998). It captures the speed with which those industry participants with above-average profits have their positive abnormal profitability revert to the industry mean. We estimate *PROFITADJ* for each industry using the industry-level panel regression over the period 1984 to 1997 that is described in the notes to Table 2.[7]

While prior literature documents a negative relation between *MATCH* and *PROFITADJ*, the interpretation of this result is unclear. Harris (1998) interprets the negative coefficient as a manifestation of proprietary costs: firms aggregate pseudo-segments from industries in which firms with above-average profits maintain their profitability advantage for long durations. However, a negative coefficient is also consistent with agency costs. That is, rather than the hidden pseudo-segments predominantly representing businesses with better profitability than their industry peers, firms may hide operations with below-average performance from shareholders because these operations are in an industry in which other firms are able to consistently achieve above-average performance.

PRIVATE is a proprietary cost proxy. Privately held U.S. firms have no obligation to publicly reveal financial results. Thus, we predict that the publicly traded firms in our sample competing in industries with high concentrations of private firms will tend to mimic the disclosure policies of their private competitors.[8] We classify a Census establishment as private if it cannot be linked to a Compustat firm. We then calculate *PRIVATE* as the ratio of total private-establishment *TVS* to total *TVS* within the pseudo-segment's industry for that Census year.[9] We predict a negative association between *PRIVATE* and the likelihood of pseudo-segment disclosure.

TRANSIN is an indicator variable equal to 1 (0) if the pseudo-segment receives (does not receive) inefficient transfers from the rest of the firm. To calculate *TRANSIN*, we adapt the procedures of Rajan et al. (2000) to our use of Census data, resulting in a three-step calculation. The first two steps produce our estimate of transfers made or received by each pseudo-segment. First, we take the difference between the pseudo-segment's actual investment and the average investment level in its four-digit SIC code, with investment defined as the ratio of net new plants (i.e., plants added less plants removed) to end-of-period plants for each five-year Census period. This is the *industry-adjusted investment ratio* (*IAIR*).

---

7 To increase the precision of our estimates, we estimate *PROFITADJ* using the entire panel of data. However, this suggests that the manager making the disclosure decision in 1987 or 1992 can predict the profit persistence in future years. Our conclusions are unchanged if we relax this assumption and estimate *PROFITADJ* using only historical data (i.e., for 1987 using 1984–1987, for 1992 using 1984–1992, and for 1997 using the entire panel).

8 A related interpretation of *PRIVATE* is that it is inversely related to barriers to entry in the industry, which may also be related to proprietary incentives to withhold information. Specifically, if industries with more privately held firms require less capital to enter (hence, the ability of private firms to stay private and avoid the capital markets), the incumbent publicly traded firms have an incentive to withhold disclosure from potential new entrants. We separately capture low barriers to entry in one of our other explanatory variables, *LOENT_HISUB*, and we find that *PRIVATE* has only a modest positive correlation with *LOENT_HISUB*. Thus, we view *PRIVATE* as primarily capturing the extent to which required disclosure by competitors is low.

9 This variable contains three potential sources of measurement error. First, the matching of Compustat observations to observations in the Census's LRD is imperfect. In particular, these matches are based on the Census's Compustat-LRD Bridge File, which contains nonpermanent Compustat identifiers taken from an edition of Compustat that is older than the one we are using. Second, some of the plants that we classify as private may be owned by non-U.S. *publicly* traded firms that make information available as part of their home-country security laws and accounting principles. Finally, we ignore sales attributable to plants located outside the U.S. when calculating the ratio.

Second, we improve the initial estimate to account for a selection effect. Specifically, the relatively large publicly traded firms in our sample likely have more funds available than the average firm included in the LRD given the inclusion of many privately held, relatively small firms in this database. The likelihood of relatively high costs of capital for the industry average firm relative to our sample firms means that the difference between a sample pseudo-segment's investment and the industry average will tend to overstate the amount of funds transferred between pseudo-segments. To correct for this selection effect, we follow Rajan et al. (2000) by subtracting from each pseudo-segment's *IAIR* the sales-weighted average of the *IAIR*s across all the pseudo-segments of the firm. This is the *industry-firm-adjusted investment ratio* (*IFAIR*).

Finally, the third step is to set *TRANSIN* equal to 0 for pseudo-segment j if: (1) $IFAIR_j$ is negative (i.e., pseudo-segment j is *making* transfers), or (2) $IFAIR_j$ is positive but the transfer received is deemed efficient. On the other hand, if $IFAIR_j$ is positive *and* the transfer received is deemed inefficient, *TRANSIN* is set equal to 1 for pseudo-segment j. We determine whether to deem a transfer as efficient or inefficient by comparing the industry Tobin's q of pseudo-segment j to the average q for all other pseudo-segments of the firm. If the transfer recipient's q is below (above) the firm average, then the transfer is deemed inefficient (efficient).[10] We predict a negative association between *TRANSIN* and the likelihood of pseudo-segment disclosure. Based on the agency cost hypothesis, we expect that firms over-investing in low-growth-opportunity pseudo-segments are less likely to separately disclose these units in their external reports.

LOENT_HISUB is an indicator variable measuring proprietary costs related to industry competition. Analytical predictions on the relation between discretionary disclosure and industry competition depend on the nature of the competition, including whether the competition is from potential entrants or incumbents and whether firms compete on setting quantity or price. Thus, one-dimensional measures of industry competition such as concentration ratios do not have clearly predictable relations to discretionary disclosure. We therefore construct *LOENT_HISUB* as a multidimensional measure of industry competition.

The *LOENT* term refers to low entry barriers, which indicate more competition from potential entrants. The *HISUB* term refers to high product substitutability, which indicates more competition from incumbents. We use the methods described in Tang (2009) to measure both phenomena. We use the ratio of capital expenditures deflated by revenues for an industry as our measure of barriers to entry arising from the need for capital investment. We measure product substitutability as the ratio of total revenue divided by the sum of raw materials and payroll costs for an industry, thus capturing the price-to-cost margin for the industry (which should be inversely related to the extent of product substitutability). Low values of the entry barrier ratio combined with high values of the product substitutability ratio indicate a pseudo-segment with high competition from both potential entrants and incumbents. Hence, we set *LOENT_HISUB* equal to 0 *unless* the former ratio is below the median value of our sample industries *and* the latter ratio is above the median value of our sample industries, in which case we set *LOENT_HISUB* equal to 1. If competition discourages discretionary disclosure, then *LOENT_HISUB* will have a negative association with pseudo-segment disclosure.

LABOR measures the ratio of total labor costs to total sales revenues for an industry. Little attention has been given to proprietary costs of disclosing information to customers or suppliers.

---

[10] We compute q ratios for 1985, 1990, and 1995 for each single-segment Compustat firm in every four-digit SIC code that has at least one Compustat segment in our sample in that year. Our q ratios are computed using the Lindenberg and Ross (1981) methodology and the specific assumptions of Hall et al. (1988). We then assign to each pseudo-segment for each five-year investment window the industry median q ratio of the Compustat single-segment firms that operate in the same industry as the pseudo-segment. We use the finest industry match that provides at least five single-segment Compustat firms with which to calculate the industry median q.

LABOR captures proprietary costs of disclosing information to suppliers of labor under the maintained assumption that these proprietary costs are greater when labor captures a larger fraction of the firm's value-added (Liberty and Zimmerman 1986; Scott 1994). If management wishes to obfuscate the true performance of the firm in order to maintain an information advantage over labor, then there will be a negative association between *LABOR* and *MATCH*.

Control Variables

Prior research indicates that greater size is generally associated with a higher level of disclosure (e.g., Lang and Lundholm 1993). We measure *FSIZE* as the natural logarithm of total firm assets (in $ millions) from Compustat.

The industry four-firm concentration ratio, *FOURFIRM*, equals the fraction of aggregate industry *TVS* that is attributable to the four industry establishments with the largest *TVS*. Higher values of *FOURFIRM* imply more industry concentration. We match the industry measure to each pseudo-segment by four-digit SIC code. Harris (1998) uses this ratio calculated from Compustat data as a measure of competition and finds that it is negatively associated with the likelihood of disclosure. Conversely, Ali et al. (2009) find no association between a Census-based concentration ratio and the decision to provide segment disclosures.

The next three control variables capture the firm's ability to aggregate the pseudo-segment within a financial statement segment. *RELSIZE* is the ratio of pseudo-segment *TVS* to firm *TVS*. We expect a higher likelihood of pseudo-segment disclosure if the pseudo-segment constitutes a bigger share of firm operations; hence, we predict a positive association between *RELSIZE* and *MATCH*. *SEGDIVERSITY* is a measure of diversity. Under SFAS No. 14, firms with operations in similar industries were afforded greater discretion to aggregate segment information. We measure segment diversity as the ratio of the number of unique two-digit SIC codes across pseudo-segments to the total number of pseudo-segments. We predict a positive association between *SEGDIVERSITY* and *MATCH*. Finally, as an additional measure of firm complexity, we include a count of the number of pseudo-segments that the firm operates in, *NUMEST*. As the count of pseudo-segments increases, aggregation of information is more likely. We thus expect a negative association between *NUMEST* and *MATCH*.

Recall that we eliminate all observations where firm-level *TVS* per the Census is either below 75 percent, or above 125 percent, of firm-level Compustat sales. This reduces measurement error, but does not eliminate it. We therefore include *CEN_CMPSTAT*, which equals the ratio of firm-level *TVS* per the Census to firm-level sales per Compustat. As this variable increases, we gain greater comfort that all corporate activity per Compustat is being captured by the Census (indicating that the firm's operations are concentrated in domestic manufacturing). Hence, there is a greater likelihood that any individual pseudo-segment will be disclosed and we expect a positive coefficient on *CEN_CMPSTAT*.

Firms report a numeric four-digit SIC code on their Census survey. However, when reporting externally, they use a textual industry description that S&P converts to a four-digit SIC code for Compustat. We use two variables that capture economic forces related to industry membership that also control for measurement error in the S&P coding. First, we calculate *INDMATCHRATE*, which equals the proportion of pseudo-segments in an industry where *MATCH* equals 1, excluding the sample pseudo-segment being analyzed. *INDMATCHRATE* captures the measurement error that results if S&P systematically misclassifies certain industries. We predict a positive association between *INDMATCHRATE* and *MATCH*.[11]

[11] This variable also captures a strategic aspect of segment disclosure. When an industry has a lower proportion of its establishments being separately disclosed as external segments, the proprietary costs of disclosing an establishment from that industry as a separate segment are likely higher. The proprietary cost hypothesis therefore also leads us to predict

We also create an indicator variable, *NEC*, for any pseudo-segment with a four-digit SIC code described as "Not Elsewhere Classified" in the SIC manual. These are potentially idiosyncratic operations where S&P might err when converting a text description to a numeric one. Even if S&P does not err, it may be easier for firms to argue that an idiosyncratic operation is not part of a separate industry, but instead should be aggregated into a broader reported segment. We predict a negative association between *NEC* and *MATCH*.

IV. RESULTS FOR MANAGERS' SEGMENT REPORTING DECISIONS

Descriptive Statistics and Sample Distribution

Panel A of Table 1 details the Compustat observations lost at each step in our sample selection and data validation processes. Note that Panels A and B of Table 1 present data at the firm level (rather than the pseudo-segment level at which our analyses are performed). Specifically, these two panels provide information on the data requirements for Compustat firms to remain in our sample (Panel A) and on the extent of within-firm disaggregation based on Census pseudo-segments versus Compustat segments (Panel B).

Panel A of Table 1 shows that we begin with 9,975 Compustat observations from the years 1987, 1992, and 1997. We lose 4,195 observations that either cannot be linked to an observation in the Census's LRD (3,072 observations) or that have incomplete Census data (1,123 observations).[12] We eliminate an additional 2,667 observations that are on both Compustat and the LRD in 1987, 1992, or 1997 because the observation does not meet our requirement of having at least two pseudo-segments. Finally, we remove 323 observations where firm-level *TVS* per the LRD exceeds 125 percent of firm-level sales per Compustat and 1,165 observations where the firm-level *TVS* is less than 75 percent of firm-level sales. These steps result in a final sample of 1,625 observations (representing 1,008 unique firms).

Table 1, Panel B contains descriptive statistics about the degree of disaggregation. The average number of Census pseudo-segments per firm, at 5.1, is more than three times the corresponding average of 1.6 Compustat segments per firm and the standard deviation of pseudo-segments per firm is more than five times the standard deviation of segments per firm. Thus, even after grouping plants by four-digit SIC code, the Census data exhibit a greater level of, and more variability in, disaggregation than the Compustat segment data.

Table 1, Panel C presents the sample's distribution by two-digit SIC code and compares it to that of the two-digit manufacturing SIC codes on Compustat.[13] The first two columns of numbers in Panel C show that pseudo-segments are fairly evenly distributed across industries, with only SIC codes 34 (Fabricated Metals and Transportation Equipment), 35 (Industrial/Commercial Machinery and Computers), and 36 (Electrical Equipment) having more than 10 percent of our sample's pseudo-segments. In contrast, *TVS* is somewhat concentrated in combined SIC codes 20

---

a positive relation between *INDMATCHRATE* and *MATCH*.

[12] Failures to match Compustat to LRD data occur for at least two reasons. First, we are using Compustat data that were matched to the LRD by Census staff on the basis of name and address. Names in the Census data may represent divisions and not ultimate parents and thus the firm may not be matched. Second, the original match was completed by Census in 2001 and CUSIP was used as the identifying variable in Compustat. If the firm changed its CUSIP (or S&P reassigned the CUSIP) after 2001, the firm will not match an observation in the edition of Compustat that we use. We compare the Compustat median sales (assets) data for matched and unmatched firms and find (in untabulated results) that the matched firms are more than three (two) times larger than unmatched firms. Matched firms have median sales (assets) of $102.5 ($82.2) million, while unmatched firms have median sales (assets) of $32.8 ($39.2) million. In addition to being smaller, we also find that the unmatched firms are less profitable (median return on sales of 2.7 percent versus 4.0 percent for the matched observations). The fact that unmatched firms are smaller corresponds to a finding reported by Maksimovic et al. (2008).

[13] Census disclosure requirements caused us to combine two industries in all situations where one of the industries exhibited fairly small representation.

TABLE 1

Comparison of Final Sample to Compustat

Panel A: Reconciliation of Compustat Sample and Final Sample

	Firm-Years
In Compustat	9,975
Less:	
Not in the Census Compustat bridge file	4,195
Has only one pseudo-segment	2,667
Aggregate *TVS* per census exceeds 125 percent of *SALES* per Compustat	323
Aggregate *TVS* per census is less than 75 percent of *SALES* per Compustat	1,165
Final Sample	1,625

Panel B: Firm Level Descriptive Statistics

Pseudo-Segments			Compustat Segments		
Mean	Median	Standard Deviation	Mean	Median	Standard Deviation
5.1	3	5.1	1.6	1	0.9

Panel C: Industry Composition

		Sample		Compustat Manufacturing		All Compustat	
Two-Digit SIC Code	Industry Description	Percentage of Pseudo-Segments	Percentage of *TVS*	Percentage of Segments	Percentage of Sales	Percentage of Segments	Percentage of Sales
20 and 21	Food and Tobacco Products	5.94	18.17	5.20	11.66	5.41	12.12
22	Textiles	3.63	3.06	1.84	0.94	1.85	0.92
23	Apparel and Other Finished Products	2.70	1.86	2.09	0.90	2.01	0.88
24	Lumber and Wood Products	3.55	2.29	1.75	1.27	1.79	1.25
25	Furniture and Fixtures	2.67	1.55	1.47	0.71	1.44	0.68

(continued on next page)

Panel C: Industry Composition

Two-Digit SIC Code	Industry Description	Sample: Percentage of Pseudo-Segments	Sample: Percentage of *TVS*	Compustat Manufacturing: Percentage of Segments	Compustat Manufacturing: Percentage of Sales	All Compustat: Percentage of Segments	All Compustat: Percentage of Sales
26	Paper and Allied Products	4.45	7.22	2.57	2.57	2.61	2.61
27	Printing and Publishing	4.42	5.92	2.83	1.38	3.69	2.03
28 and 29	Chemicals and Petroleum	5.73	6.88	14.46	20.41	14.30	20.54
30 and 31	Rubber, Plastics, and Leather	8.06	3.08	4.27	1.81	4.29	1.82
32	Stone, Clay, Glass, and Concrete	2.78	1.35	1.93	0.87	1.98	0.96
33	Primary Metals	5.76	6.42	3.85	3.84	3.87	3.81
34	Fabricated Metals and Transportation Equipment	11.34	4.73	4.86	1.51	5.00	1.55
35	Industrial/Commercial Machinery and Computers	13.33	6.26	16.65	13.28	16.38	12.81
36	Electrical Equipment	12.20	9.11	15.05	10.93	14.67	11.10
37	Transportation Equipment	5.35	16.55	5.35	22.57	5.28	21.53
38	Measuring Instruments, Photo Goods, and Watches	6.42	4.71	13.29	4.77	12.88	4.82
39	Miscellaneous	1.67	0.84	2.54	0.58	2.55	0.58
Total		100.00	100.00	100.00	100.00	100.00	100.00

Panel A provides a reconciliation of all observations on Compustat for the years 1987, 1992, and 1997 and the observations that make up our sample of 1,625 firm-years (1,008 unique firms) for the years 1987, 1992, and 1997. A firm is included in our sample if its primary SIC code is in the manufacturing sector (SIC 2000–3999) and it can be matched from Compustat to the Longitudinal Research Database (LRD) of the U.S. Census Bureau.

Panel B provides descriptive statistics for our sample.

Panel C compares the industry composition of the pseudo-segments in our sample to the industry composition of Compustat line-of-business segments. In Panel C All Compustat refers to a sample of 13,530 segments (10,442 firm-years, 5,863 firms) that have non-negative sales and assets per the Compustat database for the years 1987, 1992, or 1997, and have a primary or secondary SIC code between 2000 and 3999. Compustat Manufacturing relates to 12,358 segments (9,585 firm-years, 5,441 firms) that are a subset of All Compustat and for which 75 percent of firm-level sales is attributable to manufacturing segments (i.e., segments with either a primary or secondary four-digit SIC code between 2000 and 3999).

(continued on next page)

A *pseudo-segment* is defined as all LRD plants of a firm that operate in the same four-digit SIC code. *TVS* denotes total value of shipments for a particular pseudo-segment per the LRD. In Panel A *Aggregate TVS* is defined as follows:

$$Aggregate\ TVS = \sum_{j=1}^{n} TVS_j.$$

In the above equation j is a pseudo-segment index and n is the number of pseudo-segments for a particular firm. *SALES* denotes annual sales per Compustat (i.e., data item 12). *Compustat segments* reflect line-of-business segments reported by the firm in its external reports filed with the SEC. In Panel C the percentage of pseudo-segments represents the percentage of all sample pseudo-segments in a particular industry; percentage of *TVS* represents the percentage of the total value of shipments reported by all pseudo-segments that is attributable to the pseudo-segments of a particular industry; percentage of segments represents the percentage of all manufacturing segments in a particular industry; percentage of sales represents the percentage of all sales reported by manufacturing segments that is attributable to segments of a particular industry.

and 21 (Food and Tobacco Products), and in SIC code 37 (Transportation Equipment).

The middle two columns of numbers in Table 1, Panel C contain the industry distribution for Compustat segments reported by firms for which at least 75 percent of the firm's sales come from manufacturing industries, whereas the final two columns of the panel present the industry distribution for all Compustat segments that have primary or secondary SIC codes between 2000 and 3999. The two sets of Compustat figures are very similar and reveal that, although the industry distribution of our sample is broadly similar to that of the Compustat population, our sample is underrepresented in the Chemicals and Petroleum industry and overrepresented in the Fabricated Metals and Transportation Equipment industry.

Univariate Statistics

In Table 2, we examine how descriptive statistics for our key explanatory variables differ between the observations for which the dependent variable, *MATCH*, equals 1 and those for which it equals 0. The unit of analysis in this table, and in the remaining tables, is the pseudo-segment. We divide Table 2 into three panels that capture three different samples.

The full sample (Table 2, Panel A) includes all 1,625 firm-year observations, and thus all 8,287 of our sample pseudo-segments. The single-segment sample (Panel B) contains the 956 firm-year observations (3,300 pseudo-segments) for which the firm reports either one line-of-business segment or has no line-of-business segment footnote in its 10-K. Finally, the multi-segment sample (Panel C) contains the 669 firm-year observations (4,987 pseudo-segments) in which the firm reports two or more line-of-business segments. In each of the three samples, differences in the values of our explanatory variables emerge between the observations for which *MATCH* equals 1 and those for which it equals 0. All differences discussed below (both means and medians) are statistically significant at the 1 percent level unless otherwise noted.

For the full sample results presented in Table 2, Panel A, approximately 37 percent of the pseudo-segments are disclosed as Compustat segments (i.e., *MATCH* equals 1 for 3,032 of 8,287 observations). The median value of *I_PROFIT* is 0.08 for undisclosed pseudo-segments and 0.11 for disclosed pseudo-segments. This represents a 38 percent difference and is consistent with the agency cost motive, as it suggests that pseudo-segments are more likely to be aggregated when abnormal profitability is lower.

PROFITADJ is the *industry* rate of positive abnormal profit adjustment, where profit is based on gross margin percentage. The higher mean and median values of *PROFITADJ* for the non-disclosing sample indicate a slower speed of positive abnormal profit adjustment for the industries of non-disclosing pseudo-segments, consistent with the findings in Harris (1998), Botosan and Stanford (2005), and Berger and Hann (2007).

The 7 percent higher mean and median values of *PRIVATE* for the non-disclosing sample indicate that undisclosed pseudo-segments are from industries with higher concentrations of private firms. The mean values of *TRANSIN* indicate that 32 percent of undisclosed pseudo-segments receive inefficient transfers, whereas only 26 percent of disclosed pseudo-segments are recipients of such subsidies. Thus, consistent with the agency cost hypothesis, the univariate statistics indicate that pseudo-segments receiving subsidies are less likely to be disclosed.

LOENT_HISUB has a mean value of approximately 28 percent for both disclosed and undisclosed pseudo-segments. However, the mean and median values of *LABOR* are about 1 percent higher at undisclosed pseudo-segments, consistent with disclosure being less likely when salaries and wages represent a larger fraction of sales in the pseudo-segment's industry.

For brevity, we do not discuss the statistics related to our control variables, with the exception of *RELSIZE*, which is the ratio of pseudo-segment *TVS* to firm-level *TVS*. Untabulated results reveal that the median value of *RELSIZE* is 0.08 across the combined sample, indicating that the median pseudo-segment provides 8 percent of its firm's sales. Thus, the majority of the pseudo-

TABLE 2

Pseudo-Segment Level Descriptive Statistics

Panel A: All Sample Firms

	Mean		Median		Standard Deviation	
	MATCH = 1	*MATCH* = 0	*MATCH* = 1	*MATCH* = 0	*MATCH* = 1	*MATCH* = 0
I_PROFIT	0.15	0.05***	0.11	0.08***	0.69	0.85
PROFITADJ	0.32	0.35***	0.29	0.34***	0.39	0.37
PRIVATE	0.56	0.63***	0.58	0.65***	0.20	0.20
TRANSIN	0.26	0.32***	0.00	0.00***	0.44	0.46
LOENT_HISUB	0.28	0.28	0.00	0.00	0.45	0.45
LABOR	0.21	0.22***	0.21	0.22***	0.07	0.08
FSIZE	5.56	6.26***	5.49	6.29***	1.66	1.73
FOURFIRM	0.34	0.30***	0.31	0.26***	0.18	0.18
RELSIZE	0.36	0.10***	0.26	0.04***	0.30	0.17
SEGDIVERSITY	0.58	0.49***	0.50	0.46***	0.24	0.23
NUMEST	7.04	11.95***	5.00	8.00***	6.65	10.91
CEN_CMPSTAT	0.96	0.95***	0.96	0.94***	0.12	0.12
INDMATCHRATE	0.42	0.32***	0.41	0.33***	0.26	0.24
NEC	0.20	0.25***	0.00	0.00***	0.40	0.44
n	3,032	5,255				

Panel B: Single-Segment Firms

	Mean		Median		Standard Deviation	
	MATCH = 1	*MATCH* = 0	*MATCH* = 1	*MATCH* = 0	*MATCH* = 1	*MATCH* = 0
I_PROFIT	0.14	0.05***	0.11	0.07***	0.73	0.87
PROFITADJ	0.34	0.34	0.30	0.34	0.39	0.38
PRIVATE	0.55	0.63***	0.56	0.66***	0.20	0.20
TRANSIN	0.25	0.31***	0.00	0.00***	0.43	0.46
LOENT_HISUB	0.27	0.28	0.00	0.00	0.44	0.45
LABOR	0.20	0.21***	0.20	0.21***	0.07	0.08
FSIZE	4.99	5.61***	4.89	5.54***	1.43	1.65

(continued on next page)

Panel B: Single-Segment Firms

	Mean		Median		Standard Deviation	
	MATCH = 1	*MATCH* = 0	*MATCH* = 1	*MATCH* = 0	*MATCH* = 1	*MATCH* = 0
FOURFIRM	0.34	0.29***	0.31	0.26***	0.18	0.17
RELSIZE	0.53	0.16***	0.55	0.07***	0.31	0.21
SEGDIVERSITY	0.63	0.55***	0.50	0.50***	0.25	0.25
NUMEST	3.51	5.97***	3.00	4.00***	2.15	4.69
CEN_CMPSTAT	0.97	0.96**	0.98	0.97**	0.12	0.12
INDMATCHRATE	0.42	0.32***	0.41	0.33***	0.25	0.25
NEC	0.19	0.25***	0.00	0.00***	0.40	0.43
n	1,175	2,125				

Panel C: Multi-Segment Firms

	Mean		Median		Standard Deviation	
	MATCH = 1	*MATCH* = 0	*MATCH* = 1	*MATCH* = 0	*MATCH* = 1	*MATCH* = 0
I_PROFIT	0.15	0.05***	0.10	0.08***	0.67	0.84
PROFITADJ	0.30	0.35***	0.28	0.34***	0.39	0.36
PRIVATE	0.56	0.62***	0.59	0.65***	0.21	0.20
TRANSIN	0.27	0.32***	0.00	0.00***	0.44	0.47
LOENT_HISUB	0.29	0.28	0.00	0.00	0.45	0.45
LABOR	0.21	0.22***	0.21	0.22***	0.07	0.08
FSIZE	5.92	6.71***	5.92	6.82***	1.69	1.64
FOURFIRM	0.34	0.30***	0.31	0.26***	0.19	0.19
RELSIZE	0.25	0.06***	0.17	0.02***	0.24	0.11
SEGDIVERSITY	0.55	0.45***	0.50	0.40***	0.23	0.20
NUMEST	9.27	16.00***	7.00	13.00***	7.51	12.01
CEN_CMPSTAT	0.94	0.93***	0.94	0.92***	0.12	0.12
INDMATCHRATE	0.42	0.33***	0.41	0.33***	0.26	0.24
NEC	0.20	0.26***	0.00	0.00***	0.40	0.44
n	1,857	3,130				

*, **, *** Denote significance at the 10 percent, 5 percent, and 1 percent levels, respectively.

(continued on next page)

Panel A shows descriptive statistics for 8,287 pseudo-segments of 1,625 firm-years (1,008 unique firms) over the years 1987, 1992, and 1997.

Panel B shows descriptive statistics for 3,300 pseudo-segments of 956 firm-years over the years 1987, 1992, and 1997.

Panel C shows descriptive statistics for 4,987 pseudo-segments of 669 firm-years over the years 1987, 1992, and 1997.

The pseudo-segments described in Panel B all relate to firms that have one line-of-business segment per Compustat (i.e., single-segment firms). The pseudo-segments described in Panel C all relate to firms that have multiple line-of-business segments per Compustat (i.e., multi-segment firms).

Please refer to Table 1 for additional sample selection criteria.

A *pseudo-segment* is defined as all LRD plants of a firm that operate in the same four-digit SIC code. *MATCH* equals 1 if the pseudo-segment four-digit SIC code matches the primary or secondary SIC code of a Compustat segment, and 0 otherwise. *I_PROFIT* equals the difference between the pseudo-segment's gross margin and the industry average gross margin, divided by the standard deviation of gross margin across the industry; gross margin is obtained from the LRD database, and equals total value of shipments (*TVS*) less cost of materials less salaries and wages, with the difference scaled by *TVS*. In a regression estimated from all firms in the pseudo-segment's industry where the dependent variable is current year gross margin and the two independent variables are lagged gross margin if negative, and lagged gross margin if positive, the coefficient from the positive gross margin realizations is defined as *PROFITADJ*. *PRIVATE* equals the ratio of the sum of *TVS* across all firms in the pseudo-segment's industry that cannot be linked to Compustat (i.e., we assume that firms that are not on Compustat are privately held), divided by the sum of *TVS* across all firms in the industry. The variable *IFAIR* is used to calculate *TRANSIN*. *IFAIR* equals the proportion of a pseudo-segment's establishments that were not owned by the firm in the previous census year less the proportion of new plants for the entire industry (excluding the firm) less the weighted (by *TVS*) firm average of this difference across all of its pseudo-segments. In equation form:

$$IFAIR_j = \frac{NP_j}{EP_j} - \frac{NP_j^i}{EP_j^i} - \sum_{j=1}^{n} w_j \left(\frac{NP_j}{EP_j} - \frac{NP_j^i}{EP_j^i} \right).$$

In the equation, *NP* denotes new plants, *EP* ending plants, *w* is *RELSIZE* for the pseudo-segment, *j* indexes pseudo-segment, and *i* indexes the industry. *TRANSIN* equals 1 if *IFAIR* is positive for the pseudo-segment and the industry Tobin's Q for that pseudo-segment is less than the average of all other industry Tobin's Q measures for the firm, and 0 otherwise. *LOENT_HISUB* is an indicator variable that equals either 0 or 1. *LOENT_HISUB* equals 1 if the pseudo-segment is a member of an industry for which: (1) the ratio of industry-level capital spending to industry-level TVS in year *t* is below the year *t* median of all manufacturing industries, and (2) the ratio of industry-level *TVS* to the sum of industry-level raw materials costs and industry-level payroll costs for year *t* is above the year *t* median of all manufacturing industries. *LABOR* equals the ratio of total labor costs incurred by establishments in the same industry as the pseudo-segment to the sum of *TVS* for all establishments in the same industry as the pseudo-segment. *FOURFIRM* equals the sum of *TVS* for the four largest firms in the same industry as the pseudo-segment divided by the sum of *TVS* for all firms in the same industry as the pseudo-segment. *RELSIZE* equals the *TVS* of the pseudo-segment deflated by the total *TVS* summed across all of the firm's pseudo-segments. *SEGDIVERSITY* equals number of unique two-digit SIC codes across the firm's pseudo-segments deflated by the total number of pseudo-segments. *CEN_CMPSTAT* equals total *TVS* summed across all of the firm's pseudo-segments deflated by total sales of the firm per Compustat. *INDMATCHRATE* equals the proportion of pseudo-segments in an industry where *MATCH* equals 1, excluding the sample pseudo-segment being analyzed. *NEC* equals 1 if the four-digit SIC code contains the phrase "Not Elsewhere Classified" in the SIC manual. *NUMEST* equals the number of pseudo-segments operated by the firm. n denotes the number of observations.

segments represent a small enough portion of the firm to easily be aggregated within a financial statement segment under SFAS No. 14 rules requiring an industry segment to be separately reported if the segment's revenue, earnings, or assets are at least 10 percent of the combined total for that item across all of the firm's industry segments.[14]

Panel A of Table 2 shows that the median value of *RELSIZE* is 0.26 for disclosed pseudo-segments, versus 0.04 for those not separately disclosed as financial statement segments. This indicates that disclosed pseudo-segments generally contribute more than 10 percent of firm sales, a trigger point for separate segment disclosure under SFAS No. 14, whereas undisclosed pseudo-segments generally do not. Thus, it is critical to control for *RELSIZE* in our multivariate tests. Nevertheless, the large standard deviation for *RELSIZE* of 0.17 among undisclosed pseudo-segments indicates that a considerable portion of pseudo-segments with *RELSIZE* above 0.10 do not get disclosed as separate segments. The standard deviation of *RELSIZE* is 0.30 for disclosed pseudo-segments, which suggests that many with *RELSIZE* below 0.10 *do* get disclosed separately. Thus, *RELSIZE* is not deterministic for the pseudo-segment disclosure decision.

The results in Panels B and C of Table 2 reveal that the differences between the disclosing and non-disclosing pseudo-segments documented in Panel A also hold for the single-segment and multi-segment sample with one exception: for the single-segment sample, *PROFITADJ* is no longer statistically significantly larger for the non-disclosing pseudo-segments.

Table 3 includes a correlation matrix for the multi-segment sample with Pearson correlations above the diagonal and Spearman below.[15] Table 2 covers the main associations between our dependent variable, *MATCH*, and the treatment variables, so we do not repeat those associations here. We note from Table 3, however, that most of the explanatory variables have significant correlations not only with *MATCH*, but also with each other. We also note that *RELSIZE* has a large positive correlation with *MATCH* as well as statistically significant (although much smaller) correlations with the treatment variables. As we discuss in detail later, the relation between *RELSIZE* and *MATCH* also has an important nonlinearity due to the inclusion of a cutoff point for measures similar to *RELSIZE* in the accounting rules for segment reporting.

Base Model Logistic Regression Analysis

Full Sample

Table 4 presents the results of estimating our base model on the full, single-segment and multi-segment samples. The nine columns of numbers are divided into three sets of results for the three samples. For each, the first column of numbers presents coefficient estimates and the pseudo R^2 of the logistic regression. The pseudo R^2 exceeds 20 percent for all samples in Table 4 and in all of our subsequent logistic regressions, indicating reasonable goodness of fit. The middle column of numbers for each sample presents two-tailed p-values, and the third a measure of the economic magnitude of the variables' effects (the marginal probability effect of the variable with all other explanatory variables evaluated at their means, multiplied by one standard deviation of

---

[14] More precisely, paragraph 15 of SFAS No. 14 requires an industry segment to be classified as a reportable segment if it satisfies one or more of the following tests when the tests are applied separately for each fiscal year for which financial statements are presented: (1) Its revenue is 10 percent or more of the combined revenue of all of the enterprise's industry segments. (2) The absolute amount of its operating profit or operating loss is 10 percent or more of the greater, in absolute amount, of: (i) The combined operating profit of all industry segments that did not incur an operating loss, or (ii) The combined operating loss of all industry segments that did incur an operating loss. (3) Its identifiable assets are 10 percent or more of the combined identifiable assets of all industry segments. SFAS No. 14 goes on in paragraph 16 to state that interperiod comparability could require an industry segment to (not) be reportable even if it falls below (above) the 10 percent cutoffs in the currently reported periods.

[15] For brevity, we present correlations only for the multi-segment sample because the majority of our multivariate analyses focus on this group.

TABLE 3

Correlation Matrix for Multi-Segment Firms

	MATCH	*I_PROFIT*	*PROFITADJ*	*PRIVATE*	*TRANSIN*	*LOENT_HISUB*	*LABOR*	*FSIZE*
MATCH		0.064	−0.064	−0.142	−0.054	0.011	−0.085	−0.224
		(<0.0001)	(<0.0001)	(<0.0001)	(0.000)	(0.432)	(<0.0001)	(<0.0001)
I_PROFIT	0.056		0.029	−0.017	0.003	−0.009	−0.008	−0.031
	(<0.0001)		(0.040)	(0.235)	(0.849)	(0.526)	(0.557)	(0.028)
PROFITADJ	−0.049	0.015		0.185	−0.016	−0.072	−0.068	0.032
	(0.001)	(0.277)		(<0.0001)	(0.261)	(<0.0001)	(<0.0001)	(0.025)
PRIVATE	−0.139	−0.024	0.152		0.038	0.109	0.175	−0.133
	(<0.0001)	(0.090)	(<0.0001)		(0.008)	(<0.0001)	(<0.0001)	(<0.0001)
TRANSIN	−0.054	0.003	−0.008	0.040		0.065	0.020	0.052
	(0.000)	(0.845)	(0.567)	(0.005)		(<0.0001)	(0.154)	(0.000)
LOENT_HISUB	0.011	−0.022	−0.048	0.117	0.065		−0.139	−0.005
	(0.432)	(0.118)	(0.001)	(<0.0001)	(<0.0001)		(<0.0001)	(0.708)
LABOR	−0.083	0.009	−0.116	0.202	0.026	−0.134		−0.172
	(<0.0001)	(0.533)	(<0.0001)	(<0.0001)	(0.066)	(<0.0001)		(<0.0001)
FSIZE	−0.225	−0.027	0.037	−0.134	0.050	−0.002	−0.160	
	(<0.0001)	(0.061)	(0.008)	(<0.0001)	(0.000)	(0.889)	(<0.0001)	
FOURFIRM	0.099	0.021	0.024	−0.687	0.018	0.012	−0.252	0.177
	(<0.0001)	(0.136)	(0.084)	(<0.0001)	(0.208)	(0.413)	(<0.0001)	(<0.0001)
RELSIZE	0.536	0.108	−0.037	−0.177	−0.046	0.009	−0.066	−0.379
	(<0.0001)	(<0.0001)	(0.009)	(<0.0001)	(0.001)	(0.530)	(<0.0001)	(<0.0001)
SEGDIVERSITY	0.219	0.008	−0.006	0.028	−0.030	−0.013	0.097	−0.616
	(<0.0001)	(0.579)	(0.670)	(0.049)	(0.037)	(0.343)	(<0.0001)	(<0.0001)
NUMEST	−0.342	0.008	0.022	−0.011	0.049	−0.004	0.035	0.675
	(<0.0001)	(0.573)	(0.128)	(0.418)	(0.001)	(0.752)	(0.014)	(<0.0001)
CEN_CMPSTAT	0.051	−0.010	−0.007	0.008	−0.034	0.041	−0.038	−0.224
	(0.000)	(0.460)	(0.640)	(0.563)	(0.015)	(0.004)	(0.007)	(<0.0001)
INDMATCHRATE	0.186	0.011	−0.102	−0.327	0.023	0.046	−0.140	0.014
	(<0.0001)	(0.422)	(<0.0001)	(<0.0001)	(0.105)	(0.001)	(<0.0001)	(0.319)

(continued on next page)

	MATCH	I_PROFIT	PROFITADJ	PRIVATE	TRANSIN	LOENT_HISUB	LABOR	FSIZE
NEC	−0.064	−0.015	−0.044	0.114	−0.087	−0.133	0.091	−0.061
	(<0.0001)	(0.294)	(0.002)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)

	FOURFIRM	RELSIZE	SEGDIVERSITY	NUMEST	CEN_CMPSTAT	INDMATCHRATE	NEC
MATCH	0.093	0.465	0.232	−0.295	0.047	0.182	−0.064
	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(0.001)	(<0.0001)	(<0.0001)
I_PROFIT	0.016	0.072	0.007	0.005	0.001	−0.001	−0.030
	(0.254)	(<0.0001)	(0.616)	(0.704)	(0.924)	(0.949)	(0.032)
PROFITADJ	−0.016	−0.039	−0.004	0.000	−0.006	−0.096	−0.014
	(0.274)	(0.006)	(0.766)	(0.995)	(0.683)	(<0.0001)	(0.329)
PRIVATE	−0.705	−0.192	0.041	−0.038	0.010	−0.310	0.108
	(<0.0001)	(<0.0001)	(0.003)	(0.008)	(0.469)	(<0.0001)	(<0.0001)
TRANSIN	0.012	−0.052	−0.028	0.066	−0.035	0.022	−0.087
	(0.389)	(0.000)	(0.050)	(<0.0001)	(0.015)	(0.123)	(<0.0001)
LOENT_HISUB	0.001	0.014	−0.013	0.012	0.042	0.048	−0.133
	(0.957)	(0.311)	(0.363)	(0.382)	(0.003)	(0.001)	(<0.0001)
LABOR	−0.290	−0.069	0.115	0.001	−0.042	−0.119	0.097
	(<0.0001)	(<0.0001)	(<0.0001)	(0.938)	(0.003)	(<0.0001)	(<0.0001)
FSIZE	0.172	−0.458	−0.616	0.567	−0.228	0.018	−0.066
	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(0.194)	(<0.0001)
FOURFIRM		0.102	−0.082	0.063	−0.016	0.176	−0.279
		(<0.0001)	(<0.0001)	(<0.0001)	(0.273)	(<0.0001)	(<0.0001)
RELSIZE	0.118		0.383	−0.381	0.075	0.161	−0.065
	(<0.0001)		(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)
SEGDIVERSITY	−0.097	0.383		−0.591	0.072	0.033	0.077
	(<0.0001)	(<0.0001)		(<0.0001)	(<0.0001)	(0.019)	(<0.0001)
NUMEST	0.052	−0.540	−0.654		−0.170	−0.049	−0.007
	(0.000)	(<0.0001)	(<0.0001)		(<0.0001)	(0.001)	(0.639)

(continued on next page)

	FOURFIRM	*RELSIZE*	*SEGDIVERSITY*	*NUMEST*	*CEN_CMPSTAT*	*INDMATCHRATE*	*NEC*
CEN_CMPSTAT	−0.020	0.087	0.058	−0.167		−0.014	−0.031
	(0.166)	(<0.0001)	(<0.0001)	(<0.0001)		(0.318)	(0.028)
INDMATCHRATE	0.182	0.148	0.031	−0.028	−0.010		−0.098
	(<0.0001)	(<0.0001)	(0.029)	(0.052)	(0.486)		(<0.0001)
NEC	−0.314	−0.058	0.079	−0.011	−0.028	−0.107	
	(<0.0001)	(<0.0001)	(<0.0001)	(0.433)	(0.046)	(<0.0001)	

Correlations are estimated using 4,987 pseudo-segments from the years 1987, 1992, and 1997. The pseudo-segments all relate to firms that have multiple line-of-business segments per Compustat. Pearson (Spearman) correlations are above (below) the diagonal. p-values for two-tailed tests are shown in parentheses. Please refer to Table 1 for additional sample selection criteria and to Table 2 for variable definitions.

TABLE 4

Logistic Regression Estimates for Pseudo-Segment Disclosure Decision

		All Sample Firms			Single-Segment Firms Only			Multi-Segment Firms Only		
	Predicted Sign	Coefficient	p-value	Economic Magnitude	Coefficient	p-value	Economic Magnitude	Coefficient	p-value	Economic Magnitude
Intercept		−0.773	0.092		−1.040	0.078		−0.480	0.430	
Treatment Variables										
I_PROFIT	?	0.080	0.033	0.018	−0.020	0.716	−0.004	0.122	0.020	0.028
PROFITADJ	−	−0.139	0.128	−0.032	0.134	0.362	0.029	−0.314	0.005	−0.073
PRIVATE	−	−0.471	0.035	−0.108	−0.638	0.053	−0.140	−0.226	0.404	−0.052
TRANSIN	−	−0.157	0.004	−0.036	−0.061	−0.548	−0.013	−0.182	0.009	−0.042
LOENT_HISUB	−	0.050	0.493	0.011	0.086	0.480	0.019	−0.034	0.712	−0.008
LABOR	−	−0.675	0.171	−0.155	−0.904	−0.263	−0.198	−1.363	0.017	−0.316
Control Variables										
FSIZE		0.015	0.580	0.003	−0.091	0.023	−0.020	0.020	0.574	0.005
FOURFIRM		0.214	0.408	0.049	0.375	0.350	0.082	0.229	0.450	0.053
RELSIZE		4.073	<0.001	0.933	4.313	<0.001	0.946	5.816	<0.001	1.349
SEGDIVERSITY		0.039	0.838	−0.093	NA	NA	NA	−0.053	0.838	−0.012
NUMEST		−0.025	<0.001	−0.006	NA	NA	NA	−0.045	<0.001	−0.010
CEN_CMPSTAT		−0.407	0.178	−0.093	−0.325	0.404	−0.071	−0.207	0.550	−0.048
INDMATCHRATE		1.014	<0.001	0.232	0.779	<0.001	0.171	1.089	<0.001	0.253
NEC		−0.107	0.193	−0.024	−0.035	0.800	−0.008	−0.141	0.169	−0.032
Pseudo R^2		0.205			0.285			0.219		

The binary dependent variable is *MATCH*. All p-values are two-tailed. The "Economic Magnitude" column reflects the change in probability that the pseudo-segment is disclosed for a one standard deviation change in the variable (or an indicator variable that equals 1) with all other independent variables evaluated at their means. Please refer to Table 1 for additional sample selection criteria and to Table 2 for variable definitions. The regression depicted in the first three columns of results is estimated using 8,287 pseudo-segments from the years 1987, 1992, and 1997. The regression depicted in the middle three columns of results is estimated using 3,300 pseudo-segments from the years 1987, 1992, and 1997; the pseudo-segments all relate to firms that have a single line-of-business segment per Compustat. The regression depicted in the final three columns of results is estimated using 4,987 pseudo-segments from the years 1987, 1992, and 1997; the pseudo-segments all relate to firms that have multiple line-of-business segments per Compustat.

the variable, or by 1 if it is an indicator variable). We include multiple observations for the same firm in our regressions, so p-values are calculated using robust standard errors that correct for firm clustering.

The full sample results use both single-segment and multi-segment firms, but we argue these subsamples need to be evaluated separately. The full sample results are thus presented merely for completeness. We use them to discuss findings on the control variables, which are similar across all three samples. We defer discussion of the treatment variables for the separate single-segment and multi-segment subsample results.

The full sample results for several of the control variables are worth noting. First, the positive coefficient estimate on *RELSIZE* is consistent with pseudo-segments that represent a larger portion of their firm's sales being more likely to be disclosed. The estimate on *CEN_CMPSTAT* is not significantly different from zero. This indicates that including firm-years where Census *TVS* captures as little as 75 percent or as much as 125 percent of the total firm sales per Compustat does not distort our inferences. One variable that mitigates S&P measurement error, *INDMATCHRATE*, is highly significant, while the other, *NEC*, is not. The positive coefficient on *INDMATCHRATE* suggests that a pseudo-segment is more likely to be disclosed when other firms' pseudo-segments in the same industry tend to be disclosed. *NUMEST* captures the number of pseudo-segments at the firm level, and it is negatively associated with the likelihood of disclosing an individual pseudo-segment. This is intuitive, as more pseudo-segments increase the firm's ability to aggregate. Finally, the variable *FOURFIRM* is not related to *MATCH*.

We conclude from the control variable results that our dependent variable, *MATCH*, which captures the internal pseudo-segment disclosure aggregation decision, is reasonably associated with firm fundamentals that are likely to be non-strategic determinants of the disclosure decision.

Single-Segment Firms

The Census data are advantageous relative to Compustat data for analyzing aggregation decisions at single-segment firms. For example, Harris (1998) and Botosan and Harris (2000) evaluate only multi-segment firms. Berger and Hann (2007) and Botosan and Stanford (2005) examine firms that reported one segment under SFAS No. 14 that was restated to multiple segments under SFAS No. 131, but these restating firms represent a minority of all single-segment firms.

Single-segment firms likely differ from multi-segment firms in two related ways. First, from a strategic disclosure standpoint, line-of-business aggregation has fundamentally different consequences for single-segment versus multi-segment firms. If the firm chooses to report a single external segment, then outsiders know that *all* industries in which the firm operates are aggregated into that segment.[16] On the other hand, if the firm chooses to report multiple external segments, an additional layer of uncertainty is introduced. In this situation, there are now multiple reported divisions that might contain several possible combinations of aggregated (i.e., non-disclosed) business activities. Because this additional layer of strategy is not applicable to single-segment firms, we expect strategic factors to play a lesser role in their aggregation decisions. Second, we expect that agency costs play a smaller role in the aggregation decisions of single-segment firms because past research shows that multi-segment firms are traded at a discount relative to single-segment companies (Lang and Stulz 1994; Berger and Ofek 1995).

We present the results of our multivariate analyses of the single-segment sample in the middle three columns of numbers in Table 4.[17] The six treatment variables of interest are *I_PROFIT*,

---

[16] Outsiders will not know about the profitability of individually aggregated activities, so some uncertainty remains.
[17] In this specification, we remove the control variables *SEGDIVERSITY* and *NUMEST* because we expect that once a firm

PROFITADJ, *PRIVATE*, *TRANSIN*, *LOENT_HISUB*, and *LABOR*. For the single-segment sample, only one of the treatment variables, *PRIVATE*, is associated with the likelihood of disclosure. Hence, there is no evidence that agency costs motivate the disclosure decisions of managers of single-segment firms. Rather, proprietary costs from private competition are the key strategic determinant of pseudo-segment aggregation for these firms.

Multi-Segment Firms

The last three columns of numbers in Table 4 present the results of estimating our base model on the multi-segment firms. The overall results are considerably different from those shown in either the first three columns of numbers (i.e., the full-sample results) or the middle three columns (i.e., the single-segment sample results). For the multi-segment sample, four of the treatment variables have a significant association with *MATCH*; moreover, *PRIVATE* is not one of these variables. In contrast, *PRIVATE* is the only treatment variable that is related to the aggregation choices of single-segment firms. These results suggest that compared to single-segment firms: (1) strategic motives play a larger role in the determination of multi-segment firms' aggregation choices and (2) agency cost motives for nondisclosure are more pertinent for multi-segment firms. These results also suggest that the proprietary costs that are relevant in the multi-segment firm context are different from those relevant to managers of single-segment firms.

Turning to individual results, *I_PROFIT* has a significantly positive coefficient estimate, implying that, on average, pseudo-segments with high abnormal profitability are more likely to be disclosed. The result is consistent with the agency cost hypothesis that pseudo-segments with low abnormal profitability are more likely to be aggregated by multi-segment firms. The economic magnitude of *I_PROFIT* is also significant; a one standard deviation increase in *I_PROFIT* increases the probability of pseudo-segment disclosure by 2.8 percent.

The *PROFITADJ* variable is significantly negative, suggesting pseudo-segments are more likely to be aggregated when they operate in industries with slow convergence of abnormal profits toward the industry mean (Harris 1998). The economic magnitude column shows that, for *PROFITADJ*, a one standard deviation increase from the mean value of 0.33 is associated with a 7.3 percent decrease in disclosure probability. Harris (1998) interprets this finding as consistent with the proprietary cost motive for aggregation. Such an interpretation implies that the slow convergence of the top performers toward the industry mean is driven by the stronger firms continuing to protect their proprietary advantages with less disaggregated disclosure. An alternative possibility is that the slow convergence of the top performers toward the industry mean can be driven by the weaker firms continuing to experience unresolved agency problems and failing to fully disclose their underperforming operations (Berger and Hann 2007). We explore these alternative explanations in subsequent analyses.

The significantly negative coefficient estimate of *TRANSIN* is consistent with the agency cost hypothesis that pseudo-segments are less likely to be disclosed when they receive inefficient transfers. Again, the economic magnitude is meaningful: when the pseudo-segment is classified as receiving inefficient transfers, the probability of nondisclosure increases by 4.2 percent.

The coefficient on *LOENT_HISUB* is insignificant, inconsistent with product-market competition as reflected in low barriers to entry and high product substitutability affecting segment disaggregation. Finally, we find a negative association between the importance of labor costs in the industry and the likelihood of pseudo-segments in that industry being separately disclosed. The significantly negative coefficient estimate on *LABOR* is highly important economically, with a one

decides to report only one segment, the diversity and number of industries have no effect on the identity of the business activity that is reported as the firm's primary operation. None of our conclusions change when we include these variables in the model, and neither are significant when included.

standard deviation increase in labor costs associated with a 32 percent decrease in the probability of separately reporting the pseudo-segment.[18]

V. EXTENSIONS OF THE BASE MODEL AND SENSITIVITY ANALYSES

In this section we extend the base model estimated in Table 4. Given the sharp differences in results from estimating the base model on the multi-segment versus single-segment samples, it is clear that mixing these samples together clouds our inferences. Accordingly, we focus the remainder of the analysis on the multi-segment firms.

Nonlinear Size Effects

As discussed in footnote 14, the accounting rules for segment reporting suggest that the impact of relative pseudo-segment size on the aggregation decision is nonlinear. Specifically, the impact could depend on whether the size of the pseudo-segment exceeds the 10 percent cutoff mentioned in paragraph 15 of SFAS No. 14. Therefore, we modify the base model by adding seven new variables. In particular, we add an additional fixed effect, *SML_RELSIZE*, which is an indicator variable set equal to 1 when *RELSIZE* is less than 0.10 (i.e., a "small" pseudo-segment), and 0 otherwise (i.e., a "large" pseudo-segment). The remaining six new variables are interactions between *SML_RELSIZE* and our six treatment variables. After including these new variables, the coefficient on a particular treatment variable now reflects the relation between that variable and the likelihood that a large pseudo-segment is disclosed. The coefficients on the new interaction terms reflect how the relation varies between small and large pseudo-segments.

Table 5 reports the results from estimating this extended model.[19] Three main points emerge. First, the coefficient estimate on *SML_RELSIZE* is significantly negative, indicating that even after including the continuous variable *RELSIZE*, there remains a significant "size effect." Disclosure likelihood declines significantly when the pseudo-segment's relative size drops below the 10 percent cutoff specified by the accounting rules. Second, the coefficient estimates from the main effects of our six treatment variables are the same in Table 5 as for the multi-segment subsample in Table 4, with the exception that the estimate on *I_PROFIT* is no longer significant. Thus, pseudo-segments that have a relative size of at least 10 percent are generally affected by our treatment variables in the same manner inferred from the multi-segment results of Table 4.

Third, the coefficient estimates on the interactions between the *SML_RELSIZE* indicator and our six treatment variables are insignificant, with the exception of the estimate on *LABOR* × *SML_RELSIZE*, which is significant at the 0.10 level. We conclude that, while the disclosure decision is nonlinear with respect to pseudo-segment size (i.e., the main effect of *SML_RELSIZE* is significant), this nonlinearity does not have a first-order effect on our inferences (i.e., the interactions are almost all insignificant).

---

[18] We also estimate this model on a sample of multi-segment firms that appear in all three years of the panel. This sample consists of 2,281 pseudo-segments, or approximately 46 percent of the original sample. Our inferences are unchanged when using this specification, with the exceptions that while *I_PROFIT* and *LABOR* continue to have the same signs, the statistical significance of the estimated coefficients declines to p-values of 0.21 and 0.47, respectively.

[19] Research by Ai and Norton (2003) and Powers (2005) emphasizes that it can be potentially difficult to interpret interaction terms from nonlinear models such as probit or logistic regression models *if* one is interested in assessing marginal effects at a point other than the center of the distribution and *if* the researcher is interested in quantifying the marginal effect of the interaction term on the probability of an event (pseudo-segment disclosure in our setting). We do not believe that either of these circumstances applies to our Table 5 regression nor do they apply to the subsequent untabulated sensitivity tests that use the same interactions as in Table 5 (between *SML_RELSIZE* and our six treatment variables). Nevertheless, from an abundance of caution we further ensure that our inferences are correct by adopting the methodology developed by Norton et al. (2004) to compute marginal effects of the interaction terms and assessing these effects over the entire range of predicted probabilities from our logistic regression models. We find that our inferences are unambiguous across the entire range of predicted probabilities for all of the models that contain interactive terms.

TABLE 5

Logistic Regression Estimates for Pseudo-Segment Disclosure Decision of Multi-Segment Firms Interactions between Treatment Variables and *SML_RELSIZE* are Included

	Predicted Sign	Coefficient	p-value	Economic Magnitude
Intercept		0.605	0.347	
SML_RELSIZE		−1.558	<0.001	−0.359
Treatment Variables				
I_PROFIT	?	0.073	0.489	0.017
I_PROFIT × *SML_RELSIZE*	?	0.051	0.676	0.012
PROFITADJ	−	−0.459	0.020	−0.105
PROFITADJ × *SML_RELSIZE*	?	0.230	0.323	0.052
PRIVATE	−	−0.202	0.560	−0.046
PRIVATE × *SML_RELSIZE*	?	0.107	0.810	0.024
TRANSIN	−	−0.238	0.042	−0.053
TRANSIN × *SML_RELSIZE*	?	0.089	0.558	0.020
LOENT_HISUB	−	0.025	0.868	0.006
LOENT_HISUB × *SML_RELSIZE*	?	−0.120	0.512	−0.027
LABOR	−	−2.723	0.005	−0.620
LABOR × *SML_RELSIZE*	?	2.109	0.061	0.480
Control Variables				
FSIZE		0.035	0.346	0.008
FOURFIRM		0.245	0.418	0.056
RELSIZE		3.350	<0.001	0.763
SEGDIVERSITY		−0.008	0.975	−0.002
NUMEST		−0.043	<0.001	−0.010
CEN_CMPSTAT		−0.224	0.524	−0.051
INDMATCHRATE		1.083	<0.001	0.247
NEC		−0.140	0.181	−0.031
Pseudo R^2		0.233		

The regression is estimated using 4,987 pseudo-segments from the years 1987, 1992, and 1997. The pseudo-segments all relate to firms that have multiple line-of-business segments per Compustat. The binary dependent variable is *MATCH*. *SML_RELSIZE* is an indicator variable that equals 1 (0) if *RELSIZE* is (not) less than 0.10. All p-values are two-tailed. The "Economic Magnitude" column reflects the change in probability that the pseudo-segment is disclosed for a one standard deviation change in the variable (or an indicator variable that equals 1) with all other independent variables evaluated at their means. Please refer to Table 1 for additional sample selection criteria and to Table 2 for variable definitions.

Effects of the Excess Value of Diversification

Our results suggest that the motive and opportunity to aggregate for agency cost reasons are greater for multi-segment firms than for single-segment firms. Within the multi-segment sample, however, there is likely considerable variation in the agency cost motive. To exploit this variation, we conduct additional tests by re-estimating our base regression on two subsamples of the multi-segment sample. In particular, we separate pseudo-segments into two groups based on whether the pseudo-segment belongs to a firm with negative or nonnegative estimated excess-value from diversification. We use the methodology discussed in Berger and Ofek (1995) to estimate excess values from diversification. A negative excess value from diversification indicates that the value of the multi-segment firm is lower than the value of a collection of stand-alone firms consisting of the

multi-segment firm's separate lines of business. To the extent this value loss is a manifestation of agency problems, negative excess value firms are more likely to have agency costs than firms with nonnegative excess values.

Table 6 shows the results from estimating our base model on the subsamples of multi-segment firms that have negative and nonnegative excess values. Of the 4,987 pseudo-segments available for the logistic regressions in Tables 4 and 5, the requisite data for calculating excess values are available for 4,517. Of these, 2,545 (56 percent) pseudo-segments are from firm-years with negative excess values, and the remaining 1,972 (44 percent) pseudo-segments are from firm-years with nonnegative excess values. Overall, for all multi-segment firms, untabulated results show that the mean excess value is −0.13 and the median is −0.08, suggesting that the diversification discount identified in past literature on broad samples (Berger and Ofek 1995; Lang and Stulz 1994) exists in our sample of firms concentrated in domestic manufacturing.

TABLE 6

Logistic Regression Estimates for Pseudo-Segment Disclosure Decision of Multi-Segment Firms Separate Regressions Estimated for Firms with Negative and Nonnegative Excess Values from Diversification

		Negative Excess Value Subsample			Nonnegative Excess Value Subsample		
	Predicted Sign	Coefficient	p-value	Economic Magnitude	Coefficient	p-value	Economic Magnitude
Intercept		−0.375	0.567		−2.126	0.048	
Treatment Variables							
I_PROFIT	?	0.153	0.028	0.037	0.117	0.124	0.026
PROFITADJ	−	−0.229	0.135	−0.055	−0.399	0.007	−0.088
PRIVATE	−	−0.302	0.373	−0.072	0.113	0.794	0.025
TRANSIN	−	−0.175	0.091	−0.041	−0.201	0.087	−0.044
LOENT_HISUB	−	−0.009	0.936	−0.002	−0.026	0.873	−0.006
LABOR	−	−2.360	0.001	−0.563	0.313	0.733	0.069
Control Variables							
FSIZE		0.030	0.425	0.007	0.065	0.332	0.014
FOURFIRM		0.567	0.126	0.135	−0.091	0.858	−0.020
RELSIZE		5.142	<0.001	1.226	7.028	<0.001	1.555
SEGDIVERSITY		0.299	0.316	0.071	−0.283	0.531	−0.063
NUMEST		−0.041	<0.001	−0.010	−0.046	<0.001	−0.010
CEN_CMPSTAT		−0.402	0.290	−0.096	0.678	0.239	0.150
INDMATCHRATE		1.056	<0.001	0.252	1.266	<0.001	0.280
NEC		−0.194	0.152	−0.046	−0.042	0.779	−0.009
Pseudo R^2		0.204			0.239		

The regressions are estimated using 4,517 pseudo-segments from the years 1987, 1992, and 1997. The pseudo-segments all relate to firms that have multiple line-of-business segments per Compustat. The 2,545 pseudo-segments underlying the regression in the first three columns of results relate to firm-years in which the excess value of diversification as measured using the sales-multiple approach discussed in Berger and Ofek (1995) is negative. The 1,972 pseudo-segments underlying the regression in the second three columns of results relate to firm-years in which the excess value of diversification is non-negative. The binary dependent variable is *MATCH*. All p-values are two-tailed. The "Economic Magnitude" column reflects the change in probability that the pseudo-segment is disclosed for a one standard deviation change in the variable (or an indicator variable that equals 1) with all other independent variables evaluated at their means. Please refer to Table 1 for additional sample selection criteria and to Table 2 for variable definitions.

With regard to the proprietary cost variables, neither *PRIVATE* nor *LOENT_HISUB* are associated with *MATCH* in either the negative excess value subsample (first three columns of numbers of Table 6) or the nonnegative excess value subsample (last three columns of numbers). Hence, we continue to find no evidence that either of these factors affect multi-segment firms' pseudo-segment aggregation decisions. The results for *PROFITADJ* do, however, offer support for the proprietary cost hypothesis because the significantly negative estimate on this variable that we observe in Tables 4 and 5 is attributable to the subsample of multi-segment firms that have nonnegative excess values. Given that positive excess value firms presumably do not have an agency cost motive to suppress disclosure, this finding is consistent with *PROFITADJ* capturing a proprietary cost motive for pseudo-segment aggregation.

Turning to the agency cost variables, the estimates on *I_PROFIT* and *TRANSIN* are significant in the predicted direction for the negative excess value subsample. Thus, as expected, firms with value-reducing diversification strategies, and arguably higher agency costs, are less likely to separately disclose pseudo-segments that are less profitable, as well as those that receive more inefficient transfers. However, as shown in the final three columns of numbers of Table 6, the coefficient on *TRANSIN* is also significantly negative for the nonnegative excess value subsample. Moreover, the proprietary cost variable *LABOR* is significant in the negative excess value subsample, but not the nonnegative excess value subsample, counter to expectations. Thus, while the agency cost motive for aggregating pseudo-segments appears to apply strongly to the multi-segment firms with unsuccessful diversification strategies, there are some inconsistencies.[20]

Untabulated Sensitivity Analyses

Validation and Sensitivity Testing of our Dependent Variable, MATCH

As discussed earlier, we face a difficult research design issue in identifying the aggregation schemes firms *should* use if they neutrally apply the accounting principles that mandate separate industry disclosure. The accounting standard in effect for the sample period (SFAS No. 14) explicitly rejects applying any single industry-classification system to all firms yet, in order to conduct an objective study, we are forced to adopt a single system. In this section we provide empirical evidence supporting our earlier arguments and perform sensitivity analyses in which we use an alternative aggregation scheme based on three-digit SIC codes.

We begin by evaluating the construct validity of our disclosure measure, *MATCH*, using the data from Berger and Hann (2003, 2007), who study firms that restate historical SFAS No. 14 segment data as part of their initial adoption of SFAS No. 131. There are 99 firms that overlap between our sample and theirs.[21] We calculate two ratios at the firm level, one from each data set. First, using Berger and Hann's data, we identify the newly revealed segments in 1998 annual reports; these are segments that were not disclosed initially in 1997 under SFAS No. 14, but were subsequently revealed as restated data upon initial adoption of SFAS No. 131. For each firm, we divide these newly revealed segments by the total number of restated 1997 segments per the 1998

[20] We evaluate the robustness of the Table 6 results by estimating the extended version of the base model that incorporates *SML_RELSIZE* and its interactions with our six treatment variables. The (untabulated) results show that the only finding for the negative excess value sample sensitive to using the expanded model is that the coefficient on *I_PROFIT* is no longer significant. For the nonnegative excess value sample, two results are noteworthy. First, the marginally significant, and negative, coefficient on *TRANSIN* found in Panel B is statistically insignificant in the expanded model. Thus, the somewhat surprising finding that receiving inefficient transfers reduces the likelihood of pseudo-segment disclosure at firms with value-enhancing diversification programs is not robust. Second, similar to the results shown in Table 5, the coefficient estimate on *LABOR×SML_RELSIZE* is positive and significant.

[21] We analyze 1,625 firm years across 1987, 1992, and 1997. Berger and Hann (2003, 2007) use 796 restating firms across 1997 and 1998. The overlap is reduced from 796 because we require our firms to be almost exclusively involved in *domestic manufacturing*, whereas Berger and Hann have neither the industry nor geographic limitation *and* they sample from 1998, which we do not.

annual report. Second, using Census data, we identify the total number of unmatched pseudo-segments (i.e., *MATCH* equals 0) for each firm in 1997, and divide this by the total count of pseudo-segments for that firm-year.

Both of the ratios described above are measures of nondisclosure, and should be positively correlated. Empirically, we find that this is the case, with a Pearson correlation of 0.25 (p-value of 0.01) and a Spearman correlation of 0.22 (p-value of 0.03). Hence, our dependent variable *MATCH* exhibits external validity.

We next discuss results from using three-digit SIC codes to define industry in lieu of four-digit SIC codes. We do not believe this approach is appropriate for our primary analyses, as it adopts a disclosure benchmark that is more aggregated. If financial statement users demand detailed industry information, then a classification scheme that aggregates more and more disparate types of operations is contradictory to their demands.[22] Ultimately, this is a subjective research design choice on our part, and our conclusions based on four-digit aggregation schemes are valid only to the extent one believes, as we do, that this is a reasonable level of aggregation.

We repeat our logistic regressions for the multi-segment sample after adjusting the Table 4 specification in three ways. First, we form pseudo-segments on the basis of three-digit SIC codes and use a revised version of *MATCH* equal to 1 (0) if the pseudo-segment matches (does not match) either the primary or secondary three-digit SIC code of any of the firm's externally reported segments. Second, we define industries on the basis of three-digit SIC codes for all of our industry-adjusted (i.e., *I_PROFIT* and *TRANSIN*) and industry-level (e.g., *PROFITADJ* and *PRIVATE*) variables. Finally, we exclude *NEC*, which is designed to control for measurement error related to miscellaneous industries and/or industries that are difficult to classify. These issues are irrelevant when moving to the more aggregated three-digit SIC code level.

The (untabulated) results from the specification described above show that the treatment variables *PROFITADJ* and *TRANSIN* have strong negative and highly significant (p-values $<$ 0.01) associations with *MATCH* even when industry definitions are based on three-digit SIC codes. Hence, the main results documented in the tables are robust to an alternative industry classification scheme. However, there are two pertinent differences between the untabulated results and the results presented in Table 4 for multi-segment firms: (1) the treatment variables *I_PROFIT* and *LABOR* are no longer significant and (2) the control variable *FOURFIRM* has a positive, significant association with *MATCH*.

Effects of Information Environment Variables

As a final robustness test, we add four explanatory variables to our regressions that capture aspects of the information environment associated with disclosure quality: (1) the average number of I/B/E/S analysts covering the firm over the fiscal year; (2) return on sales; (3) a flow-of-funds based measure of the firm's net debt issuance, measured following Berger et al. (1997); and (4) a flow-of-funds based measure of new equity issued by the firm scaled by total assets [i.e., Compustat annual data items (#108)/(#6)].

The (untabulated) results from estimating this extended version of the Table 4 specification lead to the same inferences as those drawn from each subsample in Table 4. We also estimate an extended version of the Table 5 model that includes each of the four information environment variables described above. The (untabulated) results from this regression lead to the same inferences as those drawn from the results shown in Table 5, except that the coefficient estimate on *LABOR* $\times$ *SML_RELSIZE* is no longer significantly different from zero.

---

[22] The development of the more detailed NAICS suggests that even the four-digit SIC code system may be too coarse for users of industry data.

VI. CONCLUSION

We examine discretionary disclosure with unique data by examining a comprehensive set of forces hypothesized to influence segment reporting decisions. Our data consist of *confidential* information about internal operating results for a large sample of U.S. manufacturing firms, thereby capturing various aspects of agency and proprietary cost motives for nondisclosure. The combination of unique data and a broad set of disclosure determinants allows us to provide new insights regarding the motives underlying managers' aggregation decisions.

We provide evidence of differences between the disclosure behavior of single-segment firms and multi-segment firms. Single-segment firms exhibit less strategic behavior and, to the extent these firms strategize, proprietary costs related to private competition appear to be the key motive for nondisclosure. On the other hand, both proprietary and agency cost motives are important determinants of multi-segment firms' segment disclosure decisions.

One influence of agency costs on the disclosure choices of the multi-segment firms in our sample is that, if a pseudo-segment is receiving inefficient transfers of capital from the rest of the firm, it is less likely to be disclosed separately. This result is robust to alternative specifications. We also show that pseudo-segments with low industry-adjusted profits are more likely to be aggregated into another external segment and, thus, hidden from outsiders. However, this result is sensitive to model specification.

With respect to proprietary costs, we find that if a multi-segment firm has operations in an industry in which positive abnormal profits of some industry members (relative to the industry average) are likely to persist, it is less likely to disclose those operations separately. This result is robust and is manifest primarily in the subsample of firms least likely to have an agency cost motive for nondisclosure (i.e., firms for which diversification appears to be value-enhancing). This suggests a proprietary cost effect, consistent with conclusions drawn by Harris (1998). We also find that multi-segment firms are more likely to aggregate operations that are from an industry with greater labor power, although this result is sensitive to model specification.

Finally, in addition to our evidence on the proprietary and agency cost motives for aggregation, we show that comprehensive measures of industry statistics (i.e., including both privately held and publicly traded firms) are useful for measuring industry-adjusted and industry-level disclosure determinants.

REFERENCES

Ai, C., and E. Norton. 2003. Interaction terms in logit and probit models. *Economics Letters* 80 (1): 123–129.

Ali, A., S. Klasa, and E. Yeung. 2009. The limitations of industry concentration measures constructed with Compustat data: Implications for finance research. *Review of Financial Studies* 22 (10): 3839–3871.

Bamber, L., and Y. Cheon. 1998. Discretionary management earnings forecast disclosures: Antecedents and outcomes associated with forecast venue and forecast specificity choices. *Journal of Accounting Research* 36 (2): 167–190.

Bens, D., and S. Monahan. 2004. Disclosure quality and the excess value of diversification. *Journal of Accounting Research* 42 (4): 691–730.

Berger, P., and R. Hann. 2003. The impact of SFAS No. 131 on information and monitoring. *Journal of Accounting Research* 41 (2): 163–223.

———, and ———. 2007. Segment profitability and the proprietary and agency costs of disclosure. *The Accounting Review* 82 (4): 869–906.

———, and E. Ofek. 1995. Diversification's effect on firm value. *Journal of Financial Economics* 37 (1): 39–65.

———, and ———. 1999. Causes and effects of corporate refocusing programs. *Review of Financial Studies* 12 (2): 311–345.

———, ———, and D. Yermack. 1997. Managerial entrenchment and capital structure decisions. *The Journal of Finance* 52 (4): 1411–1438.

Bhojraj, S., C. Lee, and D. Oler. 2003. What's my line? A comparison of industry classification schemes for capital market research. *Journal of Accounting Research* 41 (5): 745–774.

Botosan, C., and M. Harris. 2000. Motivations for a change in disclosure frequency and its consequences: An examination of voluntary quarterly segment disclosures. *Journal of Accounting Research* 38 (2): 329–353.

——— and M. Stanford. 2005. Managers' motives to withhold segment disclosures and the effect of SFAS No. 131 on analysts' information environment. *The Accounting Review* 80 (3): 751–771.

Denis, D., D. Denis, and A. Sarin. 1997. Agency problems, equity ownership, and corporate diversification. *The Journal of Finance* 52 (1): 135–160.

Financial Accounting Standards Board (FASB). 1976. *Financial Reporting for Segments of a Business Enterprise*. Statement of Financial Accounting Standards No. 14. Norwalk, CT: FASB.

———. 1997. *Disclosures about Segments of an Enterprise and Related Information.* Statement of Financial Accounting Standards No. 131. Norwalk, CT: FASB.

Grossman, S. 1981. The role of warranties and private disclosure about product quality. *The Journal of Law and Economics* 24 (3): 461–483.

Hall, B., C. Cummins, E. Laderman, and J. Mundy. 1988. *The R&D Master File Documentation*. Boston, MA: National Bureau of Economic Research.

Harris, M. 1998. The association between competition and managers' business segment reporting decisions. *Journal of Accounting Research* 36 (1): 111–128.

Hayes, R., and R. Lundholm. 1996. Segment reporting to the capital market in the presence of a competitor. *Journal of Accounting Research* 34 (2): 261–279.

Hope, O., and W. Thomas. 2008. Managerial empire building and firm disclosure. *Journal of Accounting Research* 46 (3): 591–626.

Lamont, O. 1997. Cash flow and investment: Evidence from internal capital markets. *The Journal of Finance* 52 (1): 83–109.

Lang, L., and R. Stulz. 1994. Tobin's Q, corporate diversification and firm performance. *The Journal of Political Economy* 102 (6): 1248–1280.

Lang, M., and R. Lundholm. 1993. Cross-sectional determinants of analyst ratings of corporate disclosures. *Journal of Accounting Research* 31 (2): 246–271.

Liberty, S., and J. Zimmerman. 1986. Labor union contract negotiations and accounting choices. *The Accounting Review* 61 (4): 692–712.

Lindenberg, E., and S. Ross. 1981. Tobin's q and industrial organization. *The Journal of Business* 54 (1): 1–32.

Maksimovic, V., G. Phillips, and N. Prabhala. 2008. Post-merger restructuring and the boundaries of the firm. Working paper, University of Maryland.

McGuckin, R., and G. Pascoe. 1988. The longitudinal research database: Status and research possibilities. *Survey of Current Business* 68 (11): 30–37.

Milgrom, P. 1981. Good news and bad news: Representation theorems and applications. *The Bell Journal of Economics* 12 (2): 380–391.

Norton, E., H. Wang, and C. Ai. 2004. Computing interaction effects and standard errors in logit and probit models. *Stata Journal* 4 (2): 154–167.

Powers, E. 2005. Interpreting logit regressions with interaction terms: An application to the management turnover literature. *Journal of Corporate Finance* 11 (3): 504–522.

Rajan, R., H. Servaes, and L. Zingales. 2000. The cost of diversity: The diversification discount and inefficient investment. *The Journal of Finance* 55 (1): 35–80.

Scott, T. 1994. Incentives and disincentives for financial disclosure: Voluntary disclosure of defined benefit pension plan information by Canadian firms. *The Accounting Review* 69 (1): 26–43.

Shin, H., and R. Stulz. 1998. Are internal capital markets efficient? *The Quarterly Journal of Economics* 113 (2): 531–552.

Tang, V. 2009. Proprietary costs and privately held firms' financing choice between public offerings and private placements. Working paper, Georgetown University.

Verrecchia, R. 1983. Discretionary disclosure. *Journal of Accounting and Economics* 5 (3): 179–194.

———, and J. Weber. 2006. Redacted disclosure. *Journal of Accounting Research* 44 (4): 791–814.

THE ACCOUNTING REVIEW
Vol. 86, No. 2
2011
pp. 451–481

American Accounting Association
DOI: 10.2308/accr.00000030

The Effect of Trading Volume on Analysts' Forecast Bias

Anne Beyer
Ilan Guttman
Stanford University

ABSTRACT: This study models the interaction between a sell-side analyst and risk-averse investors. It derives an analyst's optimal earnings forecast and investors' optimal trading decisions in a setting where the analyst's payoff depends on the trading volume the forecast generates as well as on the forecast error. In the fully separating equilibrium, we find that the analyst biases the forecast upward (downward) if his private signal reveals relatively good (bad) news.

The model predicts that: (1) the analyst biases the forecast upward more often than downward and the forecast is on average optimistic; (2) the magnitude of the analyst's bias is increasing in the per-share benefit from trading volume he receives; and (3) the analyst's expected squared forecast error may increase in the precision of his private information. Finally, we characterize the circumstances under which the (rational) analyst acts as if he overweights or underweights his private information.

Keywords: *analysts' forecasts; forecast bias; trading volume; trading commission.*

I. INTRODUCTION

Sell-side analysts generate revenue for brokerage firms via trading commissions. To provide appropriate incentives, brokerage firms often tie analysts' incentives to measures of trading volume in the stocks they cover. The use of trading volume in analysts' incentive schemes is supported by Cowen et al. (2006, 125) who report that "brokerage firms usually reward their research analysts using a single measure of performance: trading volume in the stocks they cover." While trading incentives have always been part of analysts' compensation formulas, recent regulatory changes that prohibit linking analysts' compensation to investment banking activities have magnified the importance of trading incentives for analysts.

When issuing a forecast, contemporary analysts' incentives are not limited to generating trading commissions. Instead, analysts also consider their ability to generate trade in the future. As

We thank Stanley Baiman, Dean Baskin, Michael Clement, Peter DeMarzo, Ronald Dye, Paul Fischer, Steven Kachelmeier (senior editor), Robert Marquez, Paul Newman (editor), Madhav Rajan, Andrzej Skrzypacz, two anonymous referees, and participants of the 18th Annual Conference on Financial Economics and Accounting at NYU 2007, 5th Annual Conference in Corporate Finance at Olin School of Business 2008, Chicago Minnesota Theory Conference 2008, Interdisciplinary Accounting Conference 2007 in Copenhagen, Santa Clara University workshop, Stanford Accounting Summer Camp, and the Tel-Aviv Conference in Accounting for helpful comments and discussion.

Editor's note: Accepted by Paul Newman.

Submitted: July 2008
Accepted: September 2010
Published Online: March 2011

Jackson (2005, 674) states, "[t]he analyst must trade off the short-term incentive to lie and generate more trade against the long-term gains from building a good reputation." Generally, analysts are concerned about the accuracy of their forecasts because errors in forecasts can adversely affect reputation, reduce compensation (Mikhail et al. 1999), affect rankings among analysts (Stickel 1992), and call into question whether the analysts have fulfilled their fiduciary responsibility to investors (Morgan and Stocken 2003).[1]

Although forecast errors are costly to analysts, they may still choose to bias their forecasts. Prior literature has documented that analysts do in fact bias their forecasts (e.g., Dugar and Nathan 1995; Michaely and Womack 1999; Kothari 2001; Lim 2001; Hong and Kubik 2003; Jackson 2005; Chen and Jiang 2006; Bernhardt et al. 2006). While the literature studies various reasons for analysts' bias, several studies provide evidence that analysts' forecast bias is driven by trading commissions (e.g., Jackson 2005; Cowen et al. 2006).

Our objective is to develop a model of an analyst's forecasting strategy in a setting in which the analyst's payoff depends on the trading volume his forecast generates as well as on his forecast error.[2] In addition to solving for the equilibrium of the model, we derive testable empirical predictions for the bias in the analyst's forecast and its effect on the analyst's expected forecast error. While several of the model's empirical predictions are supported by existing empirical studies, others, to the best of our knowledge, have not yet been tested.

In our model, the analyst benefits from the trading volume of investors who receive his forecast (informed investors). When issuing the forecast, the analyst is not confined to tell the truth but may choose to issue a forecast that differs from his expectations. We assume that forecast errors are costly to the analyst and that the analyst trades off costs from forecast errors against incentives to generate trading volume. This trade-off determines the equilibrium properties of the analyst's forecasting strategy. In the analysis, we focus on fully separating equilibria in which investors who observe the analyst's forecast can perfectly infer the analyst's bias and private information. In a more realistic setting, one might expect that this is not the case. Our equilibrium is robust to the introduction of additional information asymmetry that prevents investors from perfectly inferring the analyst's private information (Fischer and Verrecchia 2000; Dye and Sridhar 2004).

We show that the analyst's bias is increasing in his private signal and bounded from both below and above. If the analyst's signal is sufficiently unfavorable, such that the analyst expects informed investors to sell shares, then he biases the forecast downward. Otherwise, the analyst expects informed investors to buy additional shares and biases the forecast upward. While the analyst biases the forecast both upward and downward, depending on his private signal, we show that the analyst biases the forecast upward more often than downward—resulting in the analyst's forecast being optimistic on average. This is consistent with the majority of empirical studies on analysts' bias (for a review see Kothari [2001]).

In the model, the analyst biases the forecast upward more often than downward because of investors' risk aversion. Since both informed and uninformed investors are risk-averse, their demand for the firm's shares depends on the residual uncertainty they are exposed to. The analyst's forecast decreases the uncertainty to which both informed and uninformed investors are exposed;

[1] Analysts' earnings forecast accuracy is a major factor in StarMine's evaluation of research analysts. Many Wall Street firms, including several of the ten that were involved in the Global Settlement, use StarMine data in determining their analysts' payment. Analysts' incentives may also be tied to their firms' performance through other channels such as equity compensation, bonus related to firm performance, and career concerns.

[2] For simplicity, we use male pronouns to refer to the analyst, but the gender is arbitrary; the reader should interpret each instance of "his" and "he" as meaning "his or her" and "he or she," respectively. For balance, we later arbitrarily refer to a single informed investor using the female pronouns "her" and "she."

however, the residual uncertainty of the analyst's clients (i.e., informed investors) is lower than the residual uncertainty of uniformed investors. The reason is that the information in the analyst's forecast is only partially reflected in stock prices, such that uninformed investors cannot perfectly infer the analyst's private signal. Since informed investors are exposed to less residual uncertainty than uninformed investors, the per-capita demand of informed investors for the firm's shares is on average higher than uninformed investors' per-capita demand. Therefore, the signal following which informed investors do not trade is, on average, a signal that conveys negative news. When informed investors do not trade, the analyst has no incentive to bias the forecast. Hence, in equilibrium the analyst does not bias the forecast only if his private signal equals the negative news for which informed investors do not trade on average. For more (less) favorable signals, the analyst biases the forecast upward (downward). As a result, the analyst biases the forecast upward more often than downward.

The model offers the following additional predictions. First, an increase in the analyst's per-share benefit from trading volume increases both the magnitude of the bias for any realization of the analyst's private signal and the expected forecast bias. As the analyst's per-share benefit from trading volume converges to zero, the bias converges to zero as well. This is consistent with Chen and Jiang (2006), who find that the deviation of the analyst's forecast from unbiased rational expectations increases when the analyst's benefits from doing so are high and when the costs of doing so are low.

Another prediction of the model relates the analyst's expected squared forecast error to the precision of his private information. One might expect that analysts who obtain a signal of higher precision issue a forecast that results in a smaller expected squared forecast error (distribution effect). However, the model shows that this does not have to be the case in equilibrium. Instead, the analyst's expected squared forecast error may increase in the precision of the analyst's private information. The reason is that, in addition to the distribution effect, an increase in the precision of the analyst's private signal also affects the analyst's incentives to bias the forecast. In particular, a higher precision of the analyst's private signal increases the sensitivity of informed investors' demand to the information conveyed in the analyst's forecast. This increased sensitivity boosts the analyst's incentive to bias the forecast (incentive effect). To summarize, the model predicts that analysts with higher precision of private information do not necessarily issue forecasts that result in smaller expected squared forecast errors. This result may shed some light on the surprising empirical findings that analysts who are conjectured to possess more precise information about a firm do not always outperform analysts with less precise information in terms of forecast errors (e.g., Gu and Xue 2008, in the context of affiliated versus independent analysts).

Finally, if the analyst's private signal conveys either positive or sufficiently negative news, then the analyst acts as if he overweights his private information—i.e., the analyst issues a forecast as if he places a larger weight on his private information than if he followed Bayes' Rule and truthfully reported his posterior expectation. If the analyst's private signal conveys moderately negative news, then the analyst issues a forecast as if he underweights his private information. In line with these predictions, Chen and Jiang (2006) provide evidence that analysts who issue forecasts that exceed the consensus overweight their private information. If they issue forecasts that are lower than the consensus, then analysts overweight their private information to a lesser extent and sometimes even underweight it. Easterwood and Nutt (1999), Friesen and Weller (2006), and Bernhardt et al. (2006) find that analysts do not act as if they rationally update their beliefs and truthfully report their expectations. Consistent with our model's prediction, all three studies find that analysts act as if they overweight positive private information. The empirical evidence regarding whether analysts act as if they over- or underweight negative private informa-

tion is mixed. Since our model predicts that analysts may act as if they either over- or underweight negative private information, depending on the magnitude of the negative news, our model may suggest a way to reconcile the mixed empirical evidence.

We study the analyst's forecasting decision in a setting similar to Grossman and Stiglitz (1980) wherein the analyst provides his forecast to a fraction of strategic investors. In addition to investors who observe the analyst's forecast, there are two other groups of investors: investors who are strategic in their trade but do not observe the analyst's forecast and noise traders who trade in the risky asset in a non-strategic way. This model setup guarantees that the firm's share price reflects the analyst's forecasting behavior but does not fully reveal the analyst's private information. While this market setting captures various realistic features, for example that the analyst cannot perfectly predict the trading volume that his forecast will generate, it is not designed to accommodate per-share trading costs.[3] Per-share trading costs that investors pay is an important aspect of capital markets. To study the effect of per-share trading costs that investors face, we study an alternative setting in which the analyst provides his forecast to a single investor who is small enough to act as a price-taker. This alternative setting allows us to analyze the effect of constant per-share trading costs. We find that the analyst's forecasting behavior is robust to this change in assumption and that the empirical predictions that we discussed above continue to hold.

In terms of related theoretical literature, our model is most closely related to Hayes (1998). While our model focuses on how analysts' incentives to generate trading volume affect their forecasting decision, Hayes (1998) studies how incentives to generate trading commissions affect analysts' decisions to cover stocks and gather information.[4] In contrast to Hayes (1998), we assume that the precision of the analyst's information is exogenously given, but we allow the analyst to intentionally bias the forecast. That is, the analyst may issue a forecast that deviates from his own expectation about the firm's future performance. Other related theoretical work includes Stocken (2000) and Fischer and Stocken (2010), who study the amount of information conveyed and produced by analysts in cheap-talk settings; Lim (2001) and Beyer (2008), who study analysts' forecast bias when they incur costs from forecast errors; and Guttman (2010), who endogenizes the timing of analysts' forecasts in the presence of competition between analysts. In contrast to Hayes (1998) and our model, these other models do not consider how analysts' incentives to generate trading volume affect their behavior.

II. SETUP

We model an analyst's forecasting decision when his incentives are tied to the trading volume his forecast generates. We adopt the setting of Grossman and Stiglitz (1980), where the analyst provides his forecast to a fraction $\lambda \in (0,1)$ of strategic investors ("informed investors").[5] In addition to the informed investors who receive the analyst's forecast, there are two additional groups of investors: $(1-\lambda)$ uninformed investors who are strategic in their trade but do not observe the analyst's forecast, and noise traders who trade in the risky asset in a non-strategic way. All strategic investors can allocate their wealth across a risk-free asset and a single risky asset and

[3] In order to introduce trading costs in the Grossman-Stiglitz setting and maintain the linearity of the equilibrium, we have to assume that the trading costs are quadratic. Under this assumption, the results of the model described above remain qualitatively unchanged. We pursue a different strategy in Section V of the study, since in practice marginal trading costs are unlikely to be increasing.

[4] Irvine (2004) tests empirical predictions based on Hayes (1998) and also examines how investors' trading decisions relate to the absolute value of analysts' *ex post* forecast errors.

[5] We take the fraction of informed investors as exogenous. However, the market for information is in equilibrium if informed investors have to pay the same fixed cost to obtain the private information as in Grossman and Stiglitz (1980).

exhibit homogenous CARA preferences given by $-e^{-\rho W_2}$, where W_2 is the final wealth and ρ is the risk-aversion coefficient.[6] As in Grossman and Stiglitz (1980), all strategic investors are price-takers, although in aggregate their demand affects the market price.

The market for the risky asset opens twice—once before the analyst provides the forecast to the informed investors, at $t = 0$, and once after, at $t = 1$. Grossman and Stiglitz (1980) focus on investors' equilibrium holdings at time $t = 1$. We analyze investors' equilibrium holdings not only at time $t = 1$, but also at $t = 0$ in order to derive the informed investors' trading volume at $t = 1$, which determines the analyst's payoff.

After each round of trade, the sum of the holdings of the informed investors, uninformed investors, and noise traders has to add up to 1. We assume that noise traders' holdings equal $(1 - \mu_s)$ at $t = 0$ and $(1 - s)$ at $t = 1$, respectively. The market-clearing conditions for the firm's shares at $t = 0$ and $t = 1$, respectively, are:

$$\lambda D_0^I + (1-\lambda)D_0^U = \mu_s; \tag{1}$$

$$\lambda D_1^I + (1-\lambda)D_1^U = s, \tag{2}$$

where D_t^I (D_t^U) denotes the per-capita holdings of an informed (uninformed) investor at time $t =$ 0,1. P_t denotes the market clearing price at $t = 0,1$.

We assume that noise traders' supply at $t = 1$, s, is drawn from a normal distribution with mean μ_s and variance σ_s^2.[7] Noise traders' supply at $t = 1$ is random and, hence, the equilibrium price does not fully reveal informed investors' private information. At $t = 0$, no strategic investor has private information and, hence, no random supply is needed to prevent the market price from fully revealing such information.[8] The assumption that $E[\tilde{s}] = \mu_s$ implies that the combined holdings of strategic investors at $t = 1$ are on average the same as at $t = 0$.[9] We assume $\mu_s > 0$ to reflect that strategic investors as a group hold a long-position in the firm at $t = 0$ and hold a long-position on average at $t = 1$.

The realized value of the firm's earnings, x, are not directly observable to anyone outside the firm. However, an analyst obtains private information about the firm's earnings. The analyst's private information, Ψ, is a noisy signal of firm's earnings. In particular, $\tilde{x} = \tilde{\psi} + \tilde{\varepsilon}$, where $\tilde{\psi}$ and $\tilde{\varepsilon}$ are independently and normally distributed according to $\tilde{\psi} \sim N\left(\mu_x, \frac{1}{\tau_\psi}\right)$ and $\tilde{\varepsilon} \sim N\left(0, \frac{1}{\tau_\varepsilon}\right)$, respectively. This implies that the firm's earnings are distributed normally with mean μ_x and precision $\tau_x = \frac{\tau_\psi \tau_\varepsilon}{\tau_\psi + \tau_\varepsilon}$. Based on his private information, the analyst provides a forecast, x^R, of the firm's future earnings to his clients, the informed investors.

When issuing the forecast, the analyst is not confined to tell the truth. Rather, the analyst can release a forecast that differs from his own beliefs about the firm's expected earnings given his

---

6 While we assume the same risk-aversion coefficient for all strategic investors, the characteristics of the equilibrium remain qualitatively unchanged if the risk-aversion coefficient of informed investors, $\rho_I > 0$, differs from that of uninformed investors, $\rho_U > 0$. Proofs are available upon request.

7 We follow Grossman and Stiglitz (1980) and label s as the noise traders' supply of the risky asset (even though $(1 - s)$ refers to noise traders' holdings at $t = 1$).

8 We may also allow the noise traders' supply at $t = 0$ to be random. For instance, suppose the noise traders' supply at $t = 0$, s_0, is distributed over (0,1) and noise traders' supply at $t = 1$ is normally distributed with mean s_0. As long as either the noise traders' supply s_0 or strategic investors' holdings at $t = 0$ are revealed before the analyst issues his forecast, the equilibrium remains qualitatively unchanged.

9 We could also allow for $E[\tilde{s}] > \mu_s$ without qualitatively affecting the empirical predictions. If $E[\tilde{s}] < \mu_s$, then the empirical predictions remain qualitatively the same as long as $E[\tilde{s}]$ is sufficiently close to μ_s. If $E[\tilde{s}] \ll \mu_s$, then the characteristics of the analyst's bias function remain intact except for the median bias being negative instead of positive. As a result, some empirical predictions change. For a more detailed discussion of the effect of μ_s on the analyst's reporting strategy, please see below.

private information Ψ. The analyst may want to bias the forecast in order to increase the expected trading volume generated by informed investors following the analyst's forecast.[10] However, the analyst incurs a personal cost from forecast errors where forecast errors measure the difference between the forecast and the realization of the firm's earnings. When deciding what forecast to issue, the analyst trades off the expected cost from forecast errors against the expected benefits from increased trading volume. The analyst's costs are assumed to be quadratic in the forecast error.

We assume that the analyst benefits from the trading volume his forecast generates among informed investors. Informed investors' trading volume is the absolute value of the difference between their initial holdings, λD_0^I, and their holdings after observing the analyst's forecast, λD_1^I. The noisy supply prevents the analyst from perfectly anticipating both the price P_1 and the informed investors' trade $D_1^I(x^R, P_1) - D_0^I$. Therefore, the analyst maximizes his expected benefit from trading volume net of his expected costs from forecast errors. We denote the analyst's marginal benefit from trading volume by c_A. To summarize, the analyst chooses his forecast, x^R, that maximizes his expected utility given his private information Ψ:

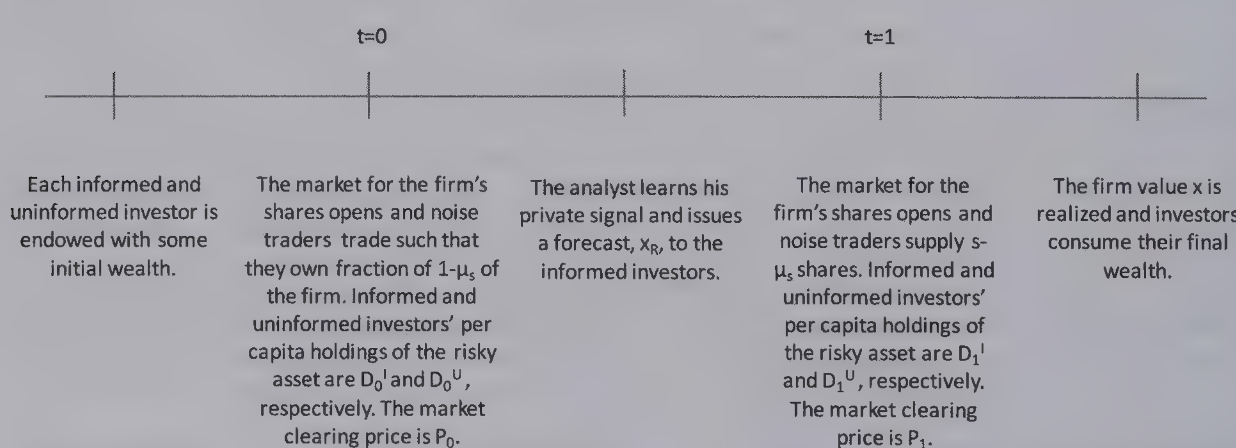
$$x^R(\psi) \in \arg\max_{x^R} c_A E[\lambda|D_1^I(\psi,\tilde{P}_1) - D_0^I||\psi] - \frac{1}{2}E[(x^R - \tilde{x})^2|\psi]. \qquad (3)$$

All parameters and the structure of the model are assumed to be common knowledge. Figure 1 summarizes the timeline of the model.

III. EQUILIBRIUM

We solve for the Perfect Bayesian Equilibrium of the game. In equilibrium, each strategic investor maximizes the expected utility of final consumption, conditional on all the information available to her including current equilibrium prices. The equilibrium price is set to clear the market. The analyst seeks to maximize the expected trading volume by informed investors fol-

FIGURE 1
Timeline



[10] One might also think of the analyst's objective function in a broader sense. In particular, his objective function can be interpreted as reflecting the extent to which his forecast changes the recipients' beliefs. Under this interpretation, the model's predictions might speak to a broader set of experts. We thank Paul Fischer for this observation.

lowing the forecast net of costs from forecast errors. When choosing his forecast, the analyst rationally anticipates informed investors' trading strategy and the trading volume it is expected to generate. The trading volume of informed investors equals the magnitude of the change in their holdings at $t = 1$ relative to their holdings at $t = 0$. In order to study the trading volume of the informed investors, we must derive the informed investors' equilibrium holdings at both $t = 1$ and $t = 0$.[11] We begin by solving for the unique linear equilibrium at $t = 1$.

Equilibrium in the Market for the Risky Asset

Three types of investors determine the equilibrium in the market for the risky asset: informed, uninformed, and noise traders. The demand of an uninformed investor at $t = 1$ is a function of the price at $t = 1$. The demand of an informed investor at $t = 1$ is a function of the price at $t = 1$ as well as of the analyst's forecast. We focus on equilibria in which the analyst's forecast fully reveals his private information, Ψ. Hence, in equilibrium, the signal informed investors infer based on the forecast they observe, $\hat{\psi}(x^R)$, equals the analyst's private signal, Ψ. The informed and uninformed investors' demand functions at $t = 1$ are the same as in Grossman and Stiglitz (1980) and are given by:

$$D_1^I(\hat{\psi}(x^R), P_1) = \frac{E(\tilde{x}|\hat{\psi}(x^R)) - P_1}{\rho Var(\tilde{x}|\hat{\psi}(x^R))},$$

$$D_1^U(P_1) = \frac{E(\tilde{x}|P_1) - P_1}{\rho Var(\tilde{x}|P_1)}.$$

Using the market-clearing condition in (2), we can derive the equilibrium price at $t = 1$. The price depends on the analyst's private signal through informed investors' demand function. Since in equilibrium the demand of informed investors is increasing in their inference about the analyst's signal, $\hat{\psi}(x^R)$, the market clearing price is also increasing in $\hat{\psi}(x^R)$. Note that the price does not fully reflect $\hat{\psi}(x^R)$ because the noisy supply garbles the signal. The higher the realization of the noisy supply, the lower is the market clearing price. In the linear equilibrium, the market clearing price can be written as:

$$P_1(\hat{\psi}(x^R), s) = \alpha_0 + \alpha_1 \hat{\psi}(x^R) - \alpha_2 s, \tag{4}$$

where $\alpha_1, \alpha_2 > 0$.

To simplify the notation, whenever our analysis is conducted on the fully revealing equilibrium path, we write both informed investors' demand and the stock price as a function of Ψ rather than a function of $\hat{\psi}(x^R)$.

In a rational expectations equilibrium, strategic investors choose equilibrium holdings at $t = 0$ anticipating the equilibrium, and their equilibrium holdings, at $t = 1$. Since strategic investors have CARA utility, their equilibrium holdings at time $t = 1$ are independent of their wealth and their prior holdings of the risky asset. As a result, at $t = 0$, both informed and uninformed investors choose the same investment strategy even though their investment strategies at $t = 1$ differ. This implies that anticipating private information in the future has no effect on today's trading strategy. In other words, investment decisions are separable in time. Since $D_0^I = D_0^U$

[11] Grossman and Stiglitz (1980) are not studying the trade generated by strategic investors, but rather focus on the equilibrium price and holdings at $t = 1$. Since the equilibrium holdings of the strategic investors at time $t = 0$ are independent of their prior investment in the risky asset due to their CARA utility function, Grossman and Stiglitz (1980) treat strategic investors' initial holdings as exogenous.

$\equiv D_0$, it follows from the market-clearing condition in (1) that the per-capita holdings of each informed and uninformed investor, D_0, equal μ_s. Lemma 1 formalizes this finding:

Lemma 1: In equilibrium, the per-capita holdings of informed and uninformed investors equal μ_s at $t = 0$.

Informed Investors' Expected Trading Volume

The analyst benefits from trading volume generated by informed investors. However, the analyst cannot perfectly predict informed investors' trading volume because volume varies with the realization of noise traders' supply. For that reason, the analyst considers informed investors' expected trading volume, $\lambda E[|D_1^I(\psi, \tilde{P}_1) - D_0^I| \mid \psi]$, when issuing the forecast. The characteristics of informed investors' expected trading volume will determine the analyst's forecasting strategy. We next derive and provide intuition for how informed investors' expected trading volume varies with the analyst's private information.

For any signal Ψ, informed investors may either buy or sell additional shares—depending on the realization of the noise traders' supply, s. Informed investors buy additional shares if the noise traders' supply is high and sell shares if it is low. Since informed investors' demand is monotonically increasing in the noise traders' supply, s, there exists a realization of noise traders' supply, s^* (ψ), such that informed investors neither buy nor sell shares:

$$s^*(\psi) = \frac{1}{\alpha_2}(\rho Var(\tilde{x}|\psi)\mu_s - E(\tilde{x}|\psi) + \alpha_0 + \alpha_1\psi). \tag{5}$$

Relying on s^* (ψ), we can write informed investors' expected trading volume, $\Delta(\Psi)$, as the sum of sell and buy transactions:

$$\Delta(\psi) \equiv \lambda E[|D_1^I(\psi,\tilde{P}_1) - D_0^I||\psi] = \lambda\int_{-\infty}^{s^*(\psi)} [D_0^I - D_1^I(\psi,P_1(\psi,s))]\phi_s(s)ds$$
$$+\lambda\int_{s^*(\psi)}^{\infty} [D_1^I(\psi,P_1(\psi,s)) - D_0^I]\phi_s(s)ds \tag{6}$$

where $\phi_s(s)$ denotes the pdf of the noise traders' supply, $\tilde{s}$. For a given Ψ, the informed investors' trading volume is V-shaped in the noise traders' supply and symmetric around the realization of the noise traders' supply for which no trade occurs, $\tilde{s} = s^*(\psi)$. Integrating the V-shaped function over the noise traders' supply, $\tilde{s}$, yields the expected trading volume informed investors generate for a given Ψ, $\Delta(\Psi)$. The signal Ψ for which this expected trading volume is minimal, must be the one for which the V-shaped function is symmetric around μ_s. We denote this signal by ψ^*. Lemma 2 characterizes the signal, ψ^*, following which informed investors' expected trading volume is minimal:

Lemma 2: Investors' expected trading volume, $\Delta(\Psi)$, is minimal when the analyst's signal equals:

$$\psi^* = \mu_x - \frac{\rho\mu_s}{\tau_\psi}.$$

Following the analyst's signal $\psi = \psi^*$, informed investors' trade is zero when $s = \mu_s$.

Lemma 2 states that informed investors' expected trading volume is minimal when the analyst's signal equals $\psi = \psi^*$. Since $\psi^* < \mu_x$, minimal expected trading volume occurs after obtaining unfavorable news. We next explore the intuition for this result.

Informed investors' expected trading volume is minimal when informed investors' holdings at $t = 1$ remain on average the same as at $t = 0$. When this is the case, uninformed investors' holdings at $t = 1$ also remain on average the same as at $t = 0$ (because noise traders on average trade zero). Moreover, per-capita holdings of informed investors at $t = 1$ will again equal those of uninformed investors because their per-capita holdings were the same at $t = 0$. Hence, we can gain intuition for the result in Lemma 2 by comparing the per-capita holdings of informed and uninformed investors at $t = 1$ for different realizations of the analyst's signal.

First, suppose the analyst's signal is $\psi = \mu_x$. In this case, the informed investors do not revise their beliefs about the firm's earnings prospects. Uninformed investors' beliefs vary with the realization of noise traders' supply, but, on average, also remain unchanged. Nevertheless, the average per-capita holdings of uninformed investors differs from that of informed investors. The reason is that both informed and uninformed investors are risk-averse and informed investors have an informational advantage over uninformed investors. As a result, informed investors are exposed to less residual uncertainty and their per-capita holdings are on average higher than those of uninformed investors when $\psi = \mu_x$.

For values of the analyst's signal other than $\psi = \mu_x$, informed and uninformed investors again hold the same beliefs about the firm's earnings prospects, on average. However, informed investors' expectations about the firm's prospects are more sensitive to a change in Ψ than uninformed investors' expectations. The reason is that uninformed investors can infer only a noisy version of Ψ from the market clearing price. We derived above that informed investors' per-capita holdings are, on average, higher than those of uninformed investors when $\psi = \mu_x$. Hence, the per-capita holdings of informed and uninformed investors are the same for an unfavorable realization of news, $\psi < \mu_x$. In particular, for $\psi = \psi^*$, the news is such that informed and uninformed investors hold the same amount of shares per capita, on average. Therefore, informed investors' expected trading volume is minimal for unfavorable news of $\psi = \psi^*$.

Lemma 3 further characterizes informed investors' expected trading volume. It shows that the expected trading volume of informed investors is a symmetric U-shaped function with its minimum at ψ^*:

Lemma 3: The informed investors' expected trading volume, $\Delta(\Psi)$, has the following properties: (i) strictly positive everywhere, (ii) convex everywhere, (iii) symmetric with respect to $\psi^* = \mu_x - \frac{\rho\mu_s}{\tau_\psi}$, and (iv) the magnitude of the slope converges to the same constant for both Ψ approaching plus and minus infinity.

The characteristics of informed investors' expected trading volume will determine the analyst's forecasting strategy. This is the case even though we consider only equilibria in which informed investors can perfectly infer the analyst's private information. The reason is that the slope of the expected trading volume will affect the analyst's marginal benefit from biasing the forecast. The following section uses the properties of informed investors' expected trading volume, established in Lemma 3, to characterize the analyst's bias function.

Characteristics of the Analyst's Bias Function

When the analyst issues a forecast, he is not confined to tell the truth. However, biasing the forecast is costly to the analyst because it increases his expected costs from forecast errors. As standard in costly signaling models, truth-telling cannot be sustained in equilibrium. The reason is as follows. If the analyst were to tell the truth, informed investors' expected trading volume would vary with the analyst's forecast. Then, the analyst's marginal expected benefit from biasing the forecast would be nonzero, while his marginal expected cost would be zero.

Before we show that there exists an equilibrium in which the analyst's forecast fully reveals his private information, we derive and provide the intuition for several properties of the analyst's

forecasting strategy. These properties are necessary for the existence of a fully separating equilibrium. First, we demonstrate that the analyst's forecasting strategy $x^R(\psi)$ has to be continuous in his private information and, hence, the analyst's equilibrium forecast must be strictly increasing in his private signal. Next, we show that not only the analyst's forecast, x^R, but also the forecast bias, $b(\psi) = x^R(\psi) - E[\tilde{x} \mid \psi]$, is increasing in the analyst's signal. Nevertheless, the bias is bounded from both below and above. Based on these characteristics, we show that there exists an equilibrium in which the analyst's forecast fully reveals his private information.

Claim 1: Following the signal ψ^*, the analyst will not bias the forecast—i.e., $x^R(\psi^*) = E[\tilde{x} \mid \psi^*]$ where $E[\tilde{x} \mid \psi^*] = \psi^* = \mu_x - \frac{\rho\mu_s}{\tau_\psi}$.

Since informed investors perfectly infer the analyst's signal, their expected trading volume is minimal following the signal ψ^* (see Lemma 2). Hence, in equilibrium, the analyst is not willing to bear any "signaling costs" in the form of biasing the forecast when he observes $\tilde{\psi} = \psi^*$.

Claim 2: The analyst's forecast $x^R(\psi)$ is continuously increasing in Ψ.

Informed investors' expected trading volume is continuous in Ψ. In equilibrium, the analyst's bias function cannot exhibit a discontinuity. If there were a discontinuity, the expected cost from forecast errors for types just to the right of the discontinuity would differ significantly from the expected cost for types just to the left, while the expected trading volume would be (almost) the same. Therefore, the type with the higher expected cost from forecast errors would always have an incentive to deviate and mimic the other type, which contradicts the existence of an equilibrium with discontinuous bias. Given that the forecast is continuous in the analyst's private information, the analyst's forecast must be strictly monotone in Ψ. This follows from the unbounded support of the analyst's private signal and the unbounded cost from forecast errors.

Claim 3: For any $\psi > \psi^*$ ($\psi < \psi^*$), the analyst's forecast bias $b(\Psi)$ is positive (negative).

Since the analyst's forecast is increasing in Ψ, informed investors infer relatively good news, i.e., $\psi > \psi^*$, from $x^R > x^R(\psi^*)$. When informed investors infer relatively good news ($\psi > \psi^*$) from the analyst's forecast and on average buy shares, the analyst biases the forecast upward to induce informed investors to buy more shares. When informed investors infer relatively bad news ($\psi < \psi^*$) from the analyst's forecast and sell shares, the analyst biases the forecast downward to induce informed investors to sell more shares and increase trading volume.[12]

Claim 4: The analyst's forecast bias $b(\Psi)$ is increasing in Ψ.

From Claims 2 and 3, it follows that the bias is continuously increasing in a neighborhood around ψ^*. Suppose the analyst's bias started to decrease for some signal to the right of ψ^*. If that were the case, then there would be two types that biased their forecasts by the same amount and, hence, incurred the same marginal expected cost from forecast errors. However, the marginal expected benefit from biasing the forecast would be different for the two types. In particular, the marginal expected benefit would be higher for the type for which the bias function is decreasing because informed investors' reaction would be more sensitive to the analyst's forecast. This implies that the analyst's bias cannot be decreasing for signals to the right of ψ^*. Based on a similar argument, we can preclude a decreasing bias for signals to the left of ψ^*.

---

[12] If all the strategic investors were informed, i.e., $\lambda = 1$, then the analyst would not bias the forecast. If $\lambda = 0$ informed investors would absorb all the exogenous supply of the noise traders, then the trading volume of informed investors would be independent of their beliefs. In turn, the analyst would minimize his expected costs by issuing an unbiased forecast because he cannot generate additional trading volume by biasing the forecast and altering informed investors' beliefs. We thank an anonymous reviewer for suggesting that we examine the effect of $\lambda = 1$ on the equilibrium.

Claim 5: The analyst's forecast bias $b(\Psi)$ is bounded from below and above.

Part (iv) of Lemma 3 establishes that the marginal expected trading volume converges to a constant. Since the expected trading volume is convex, the slope of the expected trading volume is always smaller in magnitude than its limit. Hence, the marginal benefit from biasing the forecast is bounded, which implies that the marginal expected cost must be bounded as well. The marginal expected cost can only be bounded if the analyst's bias is bounded because the cost function is infinitely steep in its tails.

While the above claims characterize the analyst's forecasting strategy for any equilibrium in which the analyst's forecast fully reveals his private information, they do not establish the existence of such equilibrium. The equilibrium exists if a solution exists to the first-order condition of the analyst's optimization problem in (3). The first-order condition yields:

$$c_A \frac{\partial E[\lambda|D_1^I(\hat{\psi},P_1(\hat{\psi},\tilde{s}))-D_0^I||\hat{\psi}]}{\partial\hat{\psi}}\frac{\partial\hat{\psi}(x^R)}{\partial x^R}-E[(x^R-\tilde{x})|\psi]=0$$

where $\hat{\psi}(x^R)$ denotes informed investors' inferences about the analyst's private signal based on the forecast he issued. In an equilibrium in which the analyst fully reveals his private information, informed investors' beliefs must be consistent with the analyst's forecasting strategy. That is, $\hat{\psi}(x^R(\psi))=\psi$. Substituting this equilibrium condition into the first-order condition yields:

$$c_A \frac{\partial E[\lambda|D_1^I(\psi,P_1(\psi,\tilde{s}))-D_0^I||\psi]}{\partial\psi}\frac{1}{\partial x^R(\psi)/\partial\psi}-E[(x^R-\tilde{x})|\psi]=0.$$

We can further compute the marginal expected trading volume (see the Appendix) and rewrite the first-order condition as:

$$c_A\lambda(1-\alpha_1)\tau_\varepsilon[1-2\Phi(s^*(\psi))]-\frac{\partial x^R(\psi)}{\partial\psi}E[(x^R-\tilde{x})|\psi]=0.$$

This differential equation for $x^R(\psi)$ has a solution with the boundary condition $x^R(\psi^*)=E[\tilde{x}\,|\,\psi^*]=\psi^*$. Based on this result and the equilibrium properties of the analyst's forecasting strategy in Claims 1 through 5, we can now establish the existence of an equilibrium in which the analyst's forecast fully reveals his private information.

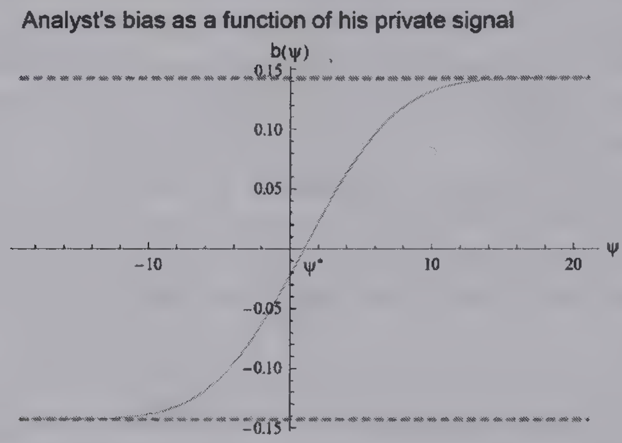
Proposition 1: There exists an equilibrium in which the analyst's forecast fully reveals his private information. In such equilibrium, the analyst's bias function $b(\Psi)$ is increasing, continuous, bounded from above and below, and $b(\psi^*)=0$.

Informed investors' beliefs are consistent with the analyst's reporting strategy in the sense that informed investors correctly infer the analyst's private information, $\hat{\psi}(x^R(\psi))=\psi$.

Figure 2 illustrates the analyst's forecasting strategy in the equilibrium of Proposition 1. To generate the figure, we used the following parameter values: $c_A=\rho=\tau_\psi=\tau_\varepsilon=\mu_s=1$, $\mu_x=2$, and $\lambda=0.5$ (which implies $\psi^*=1$).

So far, we have assumed that the analyst's payoff function is common knowledge and that the information asymmetry between the analyst and informed investors is limited to the analyst's private information, Ψ. In the equilibrium we study, the analyst's forecast fully reveals his private information. In a more realistic setting, one might expect that investors are not able to perfectly

FIGURE 2
The Analyst's Bias as a Function of His Private Signal, Ψ.



infer the analyst's private information because they are uncertain about the analyst's precise objective. The results of this study (both Proposition 1 above and Proposition 2 introduced in Section V) are robust to the introduction of such information asymmetry. More details and formal proofs are available upon request.[13]

IV. EMPIRICAL PREDICTIONS

In practice, analysts are subject to various kinds of incentives. These incentives may arise from trading volume as well as other activities such as investment banking, access to senior management of the companies covered and career concerns. Any specific set of incentives to which an analyst is exposed will affect the properties of the forecast. The empirical predictions we offer focus on the properties of analysts' bias that are due to trading volume alone.

The first prediction of the model is that the analyst biases the forecast upward more often than downward, and that his forecast is on average optimistic.

Corollary 1: The median bias and the expected bias in the analyst's forecast are positive.

The analyst's incentives to bias the forecast arise from the trading volume he expects to generate. These incentives are such that for any private signal $\psi > \psi^*$ the analyst biases the forecast upward, while for any signal $\psi < \psi^*$ the analyst biases the forecast downward. Since the median signal, $\psi = \mu_x$, is greater than $\psi^* = \mu_x - \frac{\rho\mu_s}{\tau_\psi}$, the median bias is upward and the expected bias is positive.

[13] To verify the robustness of the model to this additional information asymmetry, we introduce an additional random variable to the analyst's payoff function about which investors are uninformed (as in Fischer and Verrecchia [2000] and Dye and Sridhar [2004]). In particular, we assume that the analyst's payoff function is $u^A(x^R, \psi, \upsilon) = c_A\lambda E[|D_1^I(\psi, \tilde{P}_1) - D_0^I| \mid \psi] - E[(x^R - \bar{x} - \upsilon)^2 \mid \psi]$ where υ is normally distributed with mean zero and precision τ_υ, and is independent of x and Ψ. While the analyst observes the realization of υ before deciding about his forecast, investors do not observe it. The realization of υ may reflect idiosyncratic circumstances that influence the analysts' misreporting costs such as analysts' unknown incentives to "hype up" the firm's stock due to investment banking relations that are unobservable to outside investors. We establish the existence of an equilibrium wherein informed investors are no longer able to perfectly infer the analyst's private signal. Instead, in equilibrium informed investors can infer only the sum of the analyst's posterior expectation and the realization of υ.

The prediction that the forecast is optimistic, on average, is supported by the majority of empirical studies on analysts' earnings forecasts, as surveyed by Kothari (2001). Jackson (2005) documents that analysts trade off accuracy for optimistic bias. The literature also offers other explanations for optimistic analyst forecasts and stock recommendations. For example, Dugar and Nathan (1995), Lin and McNichols (1998), and Michaely and Womack (1999) provide evidence for affiliated analysts being more optimistic than unaffiliated analysts. Alternatively, Lim (2001) suggests that analysts rationally bias their forecast upward in order to obtain better information from the firm's management in the future. Furthermore, Hong and Kubik (2003) show that career concerns can induce overoptimistic forecasts.

In our model, both the analyst's and investors' equilibrium beliefs are fully rational in the sense that they are consistent with Bayes' Rule. Nevertheless, the equilibrium-forecasting behavior resembles the behavior of an analyst who overweights his private information for most realizations of his private signal and truthfully reports his posterior expectations. In the model, the analyst acts as if he underweights his private information only if his private information falls within the interval between $\psi^* = \mu_x - \frac{\rho\mu_s}{\tau_\psi}$ and the prior expectations of the firm's earnings μ_x. Otherwise, the analyst acts as if he overweights his private information.

Corollary 2: For all $\psi \notin [\psi^*, \mu_x]$ the analyst issues a forecast as if he overweights his private information Ψ. That is, for any $\psi \notin [\psi^*, \mu_x]$, there exists a weight $w_\psi > 1$ such that $x^R(\psi) = w_\psi \psi + (1 - w_\psi)\mu_x$. For all $\psi \in [\psi^*, \mu_x]$, the analyst issues a forecast as if he underweights his private information Ψ, i.e., for any $\psi \in [\psi^*, \mu_x]$ there exists a weight $w_\psi \leq 1$ such that $x^R(\psi) = w_\psi \psi + (1 - w_\psi)\mu_x$.

A strand of empirical literature studies the circumstances under which analysts overweight and underweight their private information when forming their forecasts. Chen and Jiang (2006) provide evidence that analysts issue forecasts as if they place larger weight on their private information than we would expect them to do if they followed Bayes' Rule and truthfully reported their posterior expectations (i.e., they act as if they overweight their private information).

Moreover, Chen and Jiang (2006) find that analysts that issue earnings forecasts that are higher than the consensus overweight their private information more than if they issue forecasts that are lower than the consensus. If analysts issue forecasts that are lower than the consensus, they sometimes even underweight their private information. These findings are consistent with the predictions of our model. In particular, our model predicts that the analyst always acts as if he overweights his private information when his private signal exceeds expectations ($\psi > \mu_x$). When his private signal is below expectations ($\psi < \mu_x$), the analyst acts as if he either over- or underweights his private information. For $\psi \in [\psi^*, \mu_x]$, the analyst underweights his private signal (i.e., $w \leq 1$). For $\psi < \psi^*$ the analyst overweights his private signal, as he does for $\psi > \mu_x$. However, the weight the analyst assigns to a negative surprise is lower than the weight he assigns to a positive surprise of the same magnitude (where the surprise is measured as $\psi - \mu_x$).

Chen and Jiang (2006) also find that the deviation from Bayesian weighting increases when the analysts' benefits from doing so are high or when the costs of doing so are low. This empirical finding is also in line with the predictions of our model, in that the analyst is fully rational and his incentives determine the extent to which the analyst acts as if he overweights/underweights his private information.

Friesen and Weller (2006) also study the bias in analysts' earnings forecasts. In particular, they contrast properties of unbiased rational forecasts and forecasts of analysts who are overconfident about the precision of their own information (cognitive bias). Their empirical findings indicate that analysts are overconfident and do not act as if they rationally update their beliefs and truthfully report their expectations. The empirical finding of Friesen and Weller (2006) is consis-

tent with our model's prediction that truth-telling cannot be sustained in equilibrium. Similar to the findings in Chen and Jiang (2006) and Friesen and Weller (2006), Easterwood and Nutt (1999) find that analysts do not act as if they rationally update their beliefs and disclose their expectations without bias. Instead, Easterwood and Nutt (1999) find that analysts over-react to positive information and under-react to negative information. In contrast to Easterwood and Nutt (1999), Bernhardt et al. (2006) find that analysts overweight both positive and negative private information. While our model predicts over-reaction to positive news and to extremely negative news, it predicts under-reaction to slightly negative news ($\psi \in [\psi^*, \mu_x]$). For positive news, our model's prediction is consistent with both Easterwood and Nutt (1999) and Bernhardt et al. (2006). For extremely negative news, our model's prediction is consistent with Bernhardt et al. (2006) but opposite to Easterwood and Nutt (1999). For slightly negative news, our model's prediction is consistent with Easterwood and Nutt (1999) but opposite to Bernhardt et al. (2006). Since both Easterwood and Nutt (1999) and Bernhardt et al. (2006) do not partition their tests according to the magnitude of the negative news, none of their findings necessarily contradicts our model's predictions. We are unaware of any empirical study that differentiates analysts' over-/under-reaction for small and large negative news.

Corollaries 1 and 2 characterized the analyst's forecasting strategy independent of the analyst's characteristics.[14] We next examine how two of the analyst's key characteristics affect the magnitude of the forecast bias and the expected squared forecast error.

In the model, the analyst's incentive to bias the forecast depends on the relative weight, c_A, his objective function assigns to the trading volume the analyst expects to generate. Higher values of c_A provide the analyst with stronger incentives to bias the forecast. As a result, the magnitude of the bias is increasing in c_A not only for each Ψ, but also in expectations. As one would expect, as c_A converges to zero, the analyst's incentive to bias the forecast diminishes and the bias converges to zero for any Ψ. The following Corollary summarizes these predictions.

Corollary 3: In the equilibrium of Proposition 1:

(a) for any $\psi \neq \psi^*$ the absolute value of $b(\tilde{\psi})$ is increasing in c_A;
(b) the expected squared forecast error, $E[FE^2]$, and the expected bias, $E[b(\tilde{\psi})]$, are increasing in c_A; and
(c) for any Ψ, $\lim_{c_A \to 0} b(\psi) = 0$.

In the model, the precision of the analyst's private information, $1 / Var(\tilde{x} \mid \psi) = \tau_\varepsilon$, is given exogenously. In practice, analysts will typically differ with respect to the precision of their private information due to various reasons, e.g., differences in ability, experience, general resources, or access to management.

One might expect that analysts whose private information is of higher precision also exhibit lower squared forecast errors on average. This would unambiguously be the case if analysts refrained from biasing their forecasts. However, because analysts bias their forecast, we need to examine the effect of the precision of the private information on the expected squared forecast error in more detail. The expected squared forecast error can be decomposed into two parts. The first part measures the remaining uncertainty and the second part captures the effect of the bias in the analyst's forecast on the expected squared forecast error:

---

[14] If uninformed investors are risk-neutral, then informed investors do not hold any shares in the firm at $t = 0$ and also on average do not hold any shares at $t = 1$. This implies that for $\tilde{\psi} = \mu_x$ informed investors on average do not change their holdings in the firm. As a result, the analyst does not bias the forecast for $\tilde{\psi} = \mu_x$ and his median and expected bias is zero. However, the properties of the analyst's bias function in Claims 1–5 still hold for $\psi^* = \mu_x$. We thank an anonymous reviewer for this observation.

$$E[FE^2] = Var(\tilde{x}|\psi) + E[b(\tilde{\psi})^2]. \tag{7}$$

When the magnitude of the analyst's bias decreases in the precision of his private information, the expected squared forecast error always decreases in the precision of the analyst's private signal. This is the case for sufficiently high precisions of the analyst's information, because informed investors' informational advantage becomes so great that uninformed investors do not trade much in the market. Hence, only informed investors absorb noise traders' supply, and their trading volume hardly varies with the analyst's forecast. As a result, the analyst has fewer incentives to bias the forecast as the precision of his private information increases.

However, when the magnitude of the analyst's bias increases in the precision of his private information, the effect of the precision of the private signal on $E[FE^2]$ is ambiguous. The magnitude of the analyst's bias increases in the precision of his private information whenever the precision is sufficiently small. Intuitively, when the precision is zero, informed investors' trading decision does not depend on the analyst's forecast and, hence, the analyst does not bias the forecast. As the precision increases, informed investors' trading decision depends on the analyst's forecast, providing him with incentives to bias the forecast. As a result, the magnitude of the analyst's bias and the expected squared forecast bias is increasing in the precision of the analyst's private information. In this case, the effect of the precision of the private signal on $E[FE^2]$ depends on the relative magnitude of two opposing effects. If the increase in the bias dominates the decrease in the conditional variance, then the expected squared forecast error increases in the precision of the analyst's private information and *vice versa*. The increase in the bias is higher when the analyst assigns a higher weight to the trading volume—i.e., for higher c_A. Hence, for sufficiently high c_A, an analyst with more precise private information produces larger expected squared forecast errors than an analyst with less precise information, as long as the precision of analysts' private information is relatively low. The following corollary summarizes the effect of the precision of the analyst's private information on the expected squared forecast error:

Corollary 4: In the equilibrium of Proposition 1, there exists c_A^* such that for all $c_A > c_A^*$ the expected squared forecast error increases in τ_ε for sufficiently small values of τ_ε.

To summarize, the model predicts that analysts with more precise private information do not necessarily issue forecasts that result in smaller expected forecast bias and/or smaller expected squared forecast errors. This might shed some light on the surprising empirical findings that affiliated analysts who are conjectured to possess more precise information about a firm do not outperform independent analysts, in the sense of issuing forecasts that result in smaller forecast errors (e.g., Gu and Xue 2008).

V. THE EFFECT OF TRADING COSTS ON THE ANALYST'S FORECAST

The previous section examined properties of the analyst's forecast when the analyst's incentives are tied to the trading volume his forecast generates among informed investors. Analysts' incentives are often tied to the trading volume in the stocks they cover because brokerage houses generate revenues from trading commissions (Cowen et al. 2006). However, in the setting studied in the previous section we assume that investors do not pay any trading commission or, more generally, do not incur any form of trading costs. We made this simplifying assumption because in a Grossman-Stiglitz-like setting in which the analyst provides his forecast to a fraction λ of strategic investors, we can only study trading strategies that are linear in the analyst's forecast. However, fixed per-share trading costs cause informed investors' trading strategy to be nonlinear in the analyst's forecast. To study the effect of per-share trading costs, we resort to a simpler setting. In particular, we assume that only one investor obtains the analyst's forecast. This investor

is sufficiently small to act as a price-taker. Apart from the investor who receives the analyst's forecast (the "informed investor"), there is a continuum of uninformed investors. As before, investors have a utility function with constant absolute risk aversion. Given these assumptions, the initial demand of the representative uninformed investor is $D_0^U = \frac{\mu_x - P_0}{\rho}\tau_x$, where $\tau_x = \frac{\tau_\psi \tau_\varepsilon}{\tau_\psi + \tau_\varepsilon}$ and P_0 is the equilibrium stock price. The exogenous per-capita supply of the firm's stock is given by μ_s.[15] In equilibrium, the firm's stock price, P_0, is set such that the per capita demand, D_0^U, equals the per-capita supply μ_s. The equilibrium stock price, that equates demand and supply, is $P_0 = \mu_x - \frac{\rho \mu_s}{\tau_x}$.

In order to capture per-share trading costs, we modify the informed investor's optimization problem. We assume the informed investor incurs per share-trading costs of c_I for each share she buys or sells, i.e., her total trading costs are $c_I|D_1^I - D_0^I|$. After observing the analyst's forecast, x^R, the informed investor updates her beliefs about the firm's earnings in a rational, Bayesian manner and chooses the demand that maximizes her expected utility (8), i.e.:

$$D_1^I(x_R) \in \arg\max_{D_1} E[u^I(D_1^I)|x^R] = -\int_{-\infty}^{\infty} e^{-\rho(W_o + D_1^I(\tilde{x} - P_0) - c_I|D_1^I - D_0^I|)} f(\tilde{x}|x^R)dx. \tag{8}$$

Anticipating her holdings following the analyst's forecast, the investor chooses her optimal initial holdings D_0^I that maximizes her expected utility.

We continue to focus on equilibria where the analyst's forecast fully reveals her private signal. In such equilibria, the informed investor's optimal holding at $t = 1$ depends on the signal that she infers from the analyst's forecast. In the absence of trading costs, the informed investor does not trade for a unique signal, $\psi = \psi^*$, for which her optimal holdings at $t = 1$ remain unchanged relative to her initial holdings D_0^I. For a range of signals close to ψ^* her optimal holdings at $t = 1$ are close to her initial holdings D_0^I, such that her marginal benefit from trading is relatively small.[16] If the informed investor incurs positive marginal trading costs, the trading costs outweight her benefits from trading for this range of signals. Hence, the informed investor does not trade for this range of signals close to ψ^*. We denote the range of signals for which the informed investor does not trade if she incurs positive marginal trading costs as the "no-trading" zone.

The informed investor chooses her holdings at $t = 0$, anticipating that she will receive the analyst's forecast. In equilibrium, the optimal initial holdings of the informed investor, D_0^I, are the same as the holdings of the representative uninformed investor, D_0^U. As a result, the fact that the informed investor anticipates receiving private information in the future has no effect on her trading strategy today. While we obtain the same result in the previous section, the result that future information receipts do not affect today's trading strategy is even more surprising here. The reason is as follows. In both models, informed investors hold more shares at $t = 1$ than uninformed investors on average, because informed investors have an informational advantage. Since informed investors hold the same amount of shares as uninformed investors initially, this implies that in both models, informed investors buy shares, on average. While trading was assumed to be free in the previous model, this model assumes that the informed investor incurs positive marginal costs from trading. Hence, we might expect the informed investor to optimally hold more shares at $t = 0$ than uninformed investors who do not trade at $t = 1$.

However, this is not the case. The intuition is as follows. The informed investor's initial holdings determine the amount of trading costs she incurs when she wants to adjust her holdings

---

[15] Since the informed investor's trade does not affect the market price, her private information is not revealed. For that reason, we do not need noise traders whose trade masks the informed investor's private information.

[16] In these kinds of models, the informed investor's certainty equivalent is quadratic in his holdings such that a small deviation from optimal holdings will also generate small marginal benefits from trading.

after observing the analyst's forecast. She can minimize her expected costs if she chooses initial holdings equal to her expected future optimal holdings. If she chooses higher initial holdings, then she will save trading costs for good news and incur additional trading costs for bad news. If she chooses lower initial holdings, then she will incur additional trading costs for good news and save trading costs for bad news. Since the informed trader is risk-averse, she chooses initial holdings that guarantee lower trading costs in low-payoff states (resulting in lower variance of final payoff) even though this increases her overall expected trading costs. The informed and uninformed investors have the same risk aversion and, hence, resolve the trade-off between higher expected payoff and lower variance in final payoff in the same way.[17] As a result, the informed trader's initial holdings are the same as those of the representative uninformed investor and the midpoint of the no-trading interval is again $\psi^* = \mu_x - \frac{\rho\mu_s}{\tau_\psi}$.

As before, the analyst maximizes the informed investor's trading volume net of his expected costs from forecast errors. In contrast to the previous section, the analyst's incentives can more readily be interpreted as maximizing the trading commission his forecast generates for the brokerage house, because the informed investor pays per-share trading commissions.[18] The analyst's objective function is similar to (3) and is given by:

$$u^A(x^R, \psi) = c_A|D_1(x^R) - D_0| - \frac{1}{2}E[(x^R - \tilde{x})|\psi]. \tag{9}$$

In a fully separating equilibrium, the analyst has no incentive to bias the forecast for all signals that fall into the "no-trading" zone because the forecast does not generate any trade. For realizations of his private signal outside the "no-trading" zone, the analyst biases the forecast to induce informed investors to buy more shares for good news and sell more shares for bad news. In fact, his forecasting strategy is qualitatively similar to the forecasting strategy derived as part of Proposition 1. In particular, the analyst's forecast bias is: (1) positive for signals for which the informed investor buys shares and negative for signals for which the informed investor sells shares; (2) weakly increasing in the analyst's private signal; and (3) bounded from above and below.[19]

Proposition 2: There exists an equilibrium in which the analyst's forecast fully reveals his private information. In this equilibrium, the analyst's bias function $b(\Psi)$ is weakly increasing, continuous, bounded from above and below, and zero for $\psi \in [\psi^* - c_I, \psi^* + c_I]$.

In particular, the analyst's equilibrium forecast is given by $x^R(\psi) = E[\tilde{x} \mid \psi] + b(\psi)$, where:[20]

$$b(\psi) = \begin{cases} \frac{c_A\tau_\varepsilon}{\rho}(1+ProductLog(-e^{-(\psi-c_I-\psi^*)\rho/c_A\tau_\varepsilon-1})) & \text{if } \psi \geq \psi^*+c_I \\ 0 & \text{if } \psi \in [\psi^*-c_I, \psi^*+c_I] \\ -\frac{c_A\tau_\varepsilon}{\rho}(1+ProductLog(-e^{-(\psi^*-c_I-\psi)\rho/c_A\tau_\varepsilon-1})) & \text{if } \psi < \psi^*-c_I. \end{cases}$$

---

17 Uninformed investors face this trade-off with respect to future expected payoffs due to their security holdings.

18 Depending on the analyst's incentive system, c_A may or may not equal c_I. While we allow for $c_I \neq c_A$, we assume that both c_I and c_A are the same for buys and sells. However, this is only a simplifying assumption and introducing different marginal costs for sells and buys will result in a similar equilibrium.

19 We omit a formal proof since it follows the same line of arguments as the proofs of Claims 1–5 preceding Proposition 1.

20 *ProductLog* (x) (also known as Lambert-W or Omega function) is the solution to the differential equation $f'(x) = \frac{f(x)}{x(1+f(x))}$. For a graph of the *ProductLog* function, see the Appendix.

The informed investor's optimal holdings at $t = 1$ are given by:

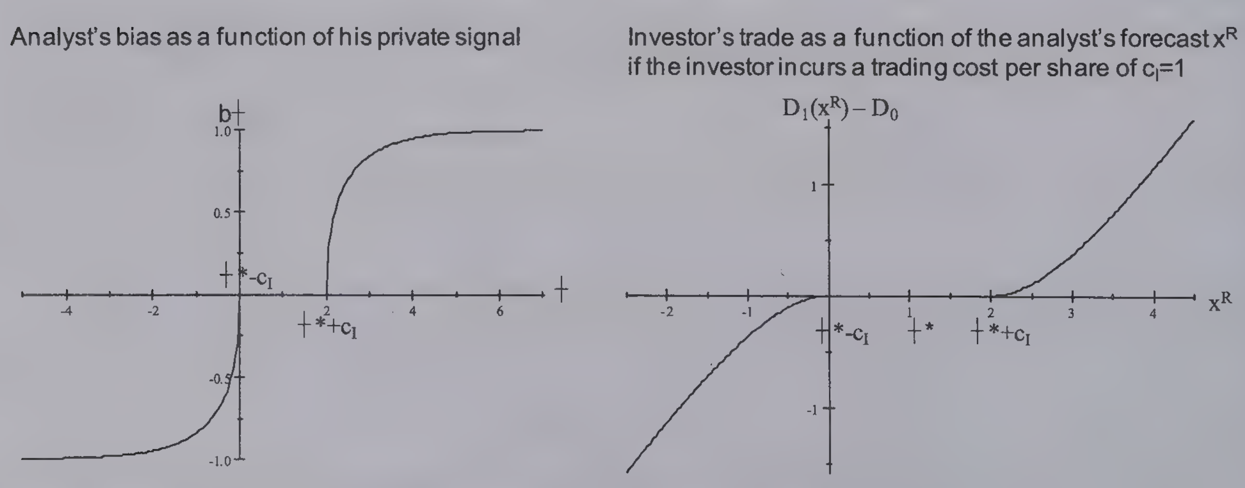
$$D_1(x^R) = \begin{cases} \dfrac{E[\tilde{x}|x^R] - P_0 - c_I}{\rho/\tau_\varepsilon} & if\ x^R > \psi^* + c_I \\ D_0 & if\ \psi^* + c_I > x^R > \psi^* - c_I \\ \dfrac{E[\tilde{x}|x^R] - P_0 + c_I}{\rho/\tau_\varepsilon} & if\ \psi^* - c_I > x^R \end{cases} \qquad (10)$$

where $P_0 = \mu_x - \frac{\rho\mu_s}{\tau_x}$. The holdings of the informed investor and the representative uninformed investor at $t = 0$ are $D_0^I = D_0^U = \mu_s$.

Figure 3 illustrates the analyst's bias function and the informed investor's demand when the informed investor faces positive marginal trading costs, c_I. For the plots, we assume the following parameter values: $c_I = c_A = \rho = \tau_\psi = \tau_\varepsilon = \mu_s = 1$; $\mu_x = 2$ (which implies $\psi^* = 1$).

The analyst's forecasting strategy in Proposition 2 is qualitatively similar to the forecasting strategy in Proposition 1. Accordingly, the empirical predictions we derive based in Proposition 1 continue to hold. In particular, even if the informed trader has to pay a per-share trading commission to the analyst, the model continues to predict the following. First, the analyst biases his forecast upward more often than downward and his forecast is on average optimistic. Second, the analyst acts as if he overweights his private information if it is favorable. If the analyst's private information is sufficiently unfavorable, then he also acts as if he overweights his private information, but to a lesser extent. If his private information is slightly unfavorable, then the analyst acts as if he underweights it. Third, the magnitude of the analyst's bias is increasing in the per-share trading commission he receives, c_A. Finally, the model shows that the analyst's expected squared forecast error does not always decrease in the precision of the analyst's information as one might expect at first. Instead, the model continues to predict that the analyst's expected squared forecast error may either increase or decrease in the precision of his private information. In particular, if the precision of his private information is sufficiently low and the per-share trading commission is sufficiently high, a further increase in this precision always increases his expected squared forecast error.

FIGURE 3
Positive Trading Costs



The only qualitative differences between the forecasting strategies in Propositions 1 and 2 are the signals for which the analyst does not bias the forecast. In the absence of per-share trading costs (Proposition 1), the analyst does not bias the forecast only when his private signal equals ψ^*. In the presence of such trading costs (Proposition 2), the analyst does not bias the forecast for an interval of private signals, which is symmetric around ψ^*. The analyst does not bias the forecast for any signal within this interval because in equilibrium, the analyst's forecast does not generate any trade by the informed investor. Since for these signals the informed investor's trading costs exceed her benefits from trading, the informed investor does not trade—eliminating the analyst's incentive to bias the forecast.

The current section adds to the analysis of the model by providing a concrete example of how analysts' payoff can be tied to the trading volume their forecasts generate. Analysts' incentives are often tied to the trading volume in the stocks they cover because brokerage houses generate revenue from trading commissions. This section analyzes how trading commissions affect investors' trading behavior and, consequently, affect analysts' forecasting strategies. In addition to motivating the assumption that analysts benefit from trading volume, this section demonstrates the robustness of the qualitative characteristics of analysts' forecasting behavior. The robustness of analysts' reporting strategy to significant model modifications, including positive marginal trading costs, suggests that the predictions of the model are likely to capture significant characteristics of analysts' forecasting behavior, whenever analysts' incentives are tied to the trading volume or the revenue from trading commissions analysts' forecasts generate for their brokerage houses.

VI. CONCLUSION

We model the interaction between sell-side analysts and their client-investors. In the model, the analyst issues a forecast to a fraction of strategic investors. The analyst benefits from the trading volume of informed investors that the forecast generates. Since the analyst is not confined to unbiased forecasts, his incentives from trading volume lead the analyst to bias the forecast. In equilibrium, the analyst trades off benefits from trading volume against the cost from forecast errors. We identify and analyze a fully separating equilibrium of this game.

The model provides several empirical predictions, including: (1) the analyst biases the forecast upward more often than downward and the forecast is optimistic, on average; (2) the magnitude of the analyst's bias is increasing in the per-share benefit from trading volume he receives; (3) the analyst's expected squared forecast error does not always decrease (as one might expect), but may also increase in the precision of his private information; and (4) the analyst acts as if he overweights his private information if it is favorable. If his private information is sufficiently unfavorable, then the analyst also acts as if he overweights it, but to a lesser extent. If his private information is slightly unfavorable, then the analyst acts as if he underweights it.

Even though part of the model's predictions are supported by existing empirical studies, the model also provides additional predictions that have not yet been tested empirically. These new predictions set the ground for future empirical work that might shed additional light on the behavior of sell-side analysts and their interaction with client-investors.

APPENDIX

Proof of Lemma 1

Investors' Demand Functions at t = 1

In a fully revealing equilibrium, an informed investor infers the analyst's private information, Ψ, from the analyst's forecast. Therefore, her demand function $D_1^I(\cdot)$ depends on the analyst's signal, Ψ, and the market equilibrium price, P_1. An informed investor's final wealth, $\tilde{W}_2^I = W_1^I + D_1^I(\tilde{x} - P_1)$, conditional on Ψ and P_1, is normally distributed according to:

$$\tilde{W}_2^I|\psi,P_1 \sim N(W_1^I + (E(\tilde{x}|\psi) - P_1)D_1^I,(D_1^I)^2Var(\tilde{x}|\psi)),$$

where W_1^I is the informed investor's wealth at $t = 1$.[21] The informed investor's expected utility at $t = 1$ is given by:

$$E[u^I(D_1^I)|\psi,P_1] = E[e^{-\rho\tilde{W}_2^I}|\psi,P_1] = -e^{-\rho(W_1^I + D_1^I(E(\tilde{x}|\psi)-P_1)-\rho/2(D_1^I)^2Var(\tilde{x}|\psi))}.$$

The demand that maximizes the informed investor's expected utility is:

$$D_1^I(\psi,P_1) = \frac{E(\tilde{x}|\psi) - P_1}{\rho Var(\tilde{x}|\psi)}.$$

Next, we solve for an uninformed investor's demand function. As in Grossman and Stiglitz (1980), we conjecture an equilibrium in which the pricing function at $t = 1$ is linear in both the noise traders' supply, s, and the analyst's private signal, Ψ:

$$P_1(\psi,s) = \alpha_0 + \alpha_1\psi - \alpha_2 s. \tag{11}$$

Since $\tilde{x}$ and $\tilde{P}_1$ are jointly normally distributed, the expected utility of the uninformed trader (given the price P_1) is given by:

$$E[u^U(D_1^U)|P_1] = E[e^{-\rho\tilde{W}_2^U}|P_1] = -e^{-\rho(W_1^U + (E(\tilde{x}|P_1)-P_1)D_1^U-\rho/2(D_1^U)^2Var(\tilde{x}|P_1))}.$$

The demand that maximizes the uninformed investor's expected utility is:

$$D_1^U(P_1) = \frac{E(\tilde{x}|P_1) - P_1}{\rho Var(\tilde{x}|P_1)}.$$

Equilibrium Price at t = 0

In order to solve for $P_1(\psi, s)$, we need to compute $E(\tilde{x} \mid P_1)$. P_1 is informationally equivalent to $\frac{\lambda E(\tilde{x}|\psi)}{\rho Var(\tilde{x}|\psi)} - s$ or $\psi - \frac{\rho}{\lambda\tau_\varepsilon}s$. Hence:

$$E(\tilde{x}|P_1) = E\left(\tilde{x}\middle|\psi - \frac{\rho}{\lambda\tau_\varepsilon}s\right) = \frac{\psi + \tau_\psi\left(\frac{\rho}{\lambda\tau_\varepsilon}\right)^2\sigma_s^2\mu_x - \frac{\rho}{\lambda\tau_\varepsilon}(s-\mu_s)}{1 + \tau_\psi\left(\frac{\rho}{\lambda\tau_\varepsilon}\right)^2\sigma_s^2}.$$

Solving the market-clearing condition in (2) for $P_1(\psi, s)$ yields:

$$P_1(\psi,s) = \frac{\frac{\lambda E(\tilde{x}|\psi)}{Var(\tilde{x}|\psi)} + \frac{(1-\lambda)E(\tilde{x}|P_1)}{Var(\tilde{x}|P_1)} - \rho s}{\frac{\lambda}{Var(\tilde{x}|\psi)} + \frac{(1-\lambda)}{Var(\tilde{x}|P_1)}}, \tag{12}$$

where:

$$E(\tilde{x}|\psi) = \psi;$$

[21] W_1^I is the informed investor's wealth at $t = 1$, i.e., $W_1^I = M_0^I + D_0^I P_1 = M_1^I + D_1^I P_1$ where M_t^I is her investment in the risk-free asset at $t = 0,1$.

$$Var(\tilde{x}|\psi) = \frac{1}{\tau_\varepsilon}; \text{ and}$$

$$Var(\tilde{x}|P_1) = Var\left(\tilde{x}\Big|\psi - \frac{\rho}{\lambda\tau_\varepsilon}s\right) = \frac{1 + (\tau_\psi + \tau_\varepsilon)\left(\frac{\rho}{\lambda\tau_\varepsilon}\right)^2\sigma_s^2}{\tau_\varepsilon\left(1 + \tau_\psi\left(\frac{\rho}{\lambda\tau_\varepsilon}\right)^2\sigma_s^2\right)}.$$

Substituting into (12) yields:

$$P_1(\psi,s) = \frac{\left(\lambda(\tau_\psi + \tau_\varepsilon)\left(\frac{\rho}{\lambda\tau_\varepsilon}\right)^2\sigma_s^2 + 1\right)\left(\psi - \frac{\rho}{\lambda\tau_\varepsilon}s\right) + (1-\lambda)\left(\tau_\psi\left(\frac{\rho}{\lambda\tau_\varepsilon}\right)^2\sigma_s^2\mu_x + \frac{\rho}{\lambda\tau_\varepsilon}\mu_s\right)}{(\tau_\psi + \lambda\tau_\varepsilon)\left(\frac{\rho}{\lambda\tau_\varepsilon}\right)^2\sigma_s^2 + 1},$$

and solving for the coefficients of the pricing function $P_1(\psi, s)$ in (11) yields:

$$\alpha_1 = \frac{\lambda(\tau_\psi + \tau_\varepsilon)\left(\frac{\rho}{\lambda\tau_\varepsilon}\right)^2\sigma_s^2 + 1}{(\tau_\psi + \lambda\tau_\varepsilon)\left(\frac{\rho}{\lambda\tau_\varepsilon}\right)^2\sigma_s^2 + 1};$$

$$\alpha_2 = \alpha_1\frac{\rho}{\lambda\tau_\varepsilon};$$

$$\alpha_0 = \frac{(1-\lambda)\left(\tau_\psi\left(\frac{\rho}{\lambda\tau_\varepsilon}\right)^2\sigma_s^2\mu_x + \frac{\rho}{\lambda\tau_\varepsilon}\mu_s\right)}{(\tau_\psi + \lambda\tau_\varepsilon)\left(\frac{\rho}{\lambda\tau_\varepsilon}\right)^2\sigma_s^2 + 1}. \tag{13}$$

Investors' Demand Functions at $t = 0$

We start by computing the expected utility of both an informed and uninformed investor conditional on P_1. The expected utility of the uninformed investor is:

$$E[u^U(D_1^U)|P_1] = -\ e^{-\rho(W_1^U + D_1^U(E(\tilde{x}|P_1) - P_1) - \rho/2(D_1^U)^2 Var(\tilde{x}|P_1))}\Big|_{D_1^U(P_1) = E(\tilde{x}|P_1) - P_1/\rho Var(\tilde{x}|P_1)} = \\ -e^{-\rho W_1^U}e^{-(E(\tilde{x}|P_1)\ -\ P_1)^2/2Var(\tilde{x}|P_1)},$$

where $W_1^U = W_0^U + D_0^U(P_1 - P_0)$. The expected utility of the informed investor conditional on both P_1 and Ψ is:

$$E[u^I(D_1^I)|\psi, P_1] = -\ e^{-\rho(W_1^I + D_1^I(E(\tilde{x}|\psi) - P_1) - \rho/2(D_1^I)^2 Var(\tilde{x}|\psi))}\Big|_{D_1^I(\psi,P_1) = E(\tilde{x}|\psi) - P_1/\rho Var(\tilde{x}|\psi)} = \\ -e^{-\rho W_1^I}e^{-(E(\tilde{x}|\psi)\ -\ P_1)^2/2Var(\tilde{x}|\psi)},$$

where $W_1^I = W_0^I + D_0^I(P_1 - P_0)$. Since $\tilde{\psi}|\,P_1$ is normally distributed, we can define:

$$\tilde{Z} = \left. \frac{E(\tilde{x}|\psi) - P_1}{\sqrt{Var(E(\tilde{x}|\psi)|P_1)}} \right|_{P_1},$$

where $\tilde{Z}$ is distributed normally with mean $\mu_z = \frac{E(\tilde{x}|P_1)-P_1}{\sqrt{Var(E(\tilde{x}|\psi)|P_1)}}$ and variance 1. $\tilde{Z}^2$ is distributed non-central Chi-square with one degree of freedom and parameter $k = \mu_z^2$. The moment-generating function for the non-central Chi-square distribution is $E[e^{-t\tilde{Z}^2}] = \frac{1}{\sqrt{1+2t}} \exp\{-\frac{t}{1+2t}\mu_z^2\}$, where $t = \frac{Var(E(\tilde{x}|\psi)|P_1)}{2Var(\tilde{x}|\psi)}$. Simplification yields;

$$E[e^{-t\tilde{Z}^2}] = \frac{1}{\sqrt{1 + 2\frac{Var(E(\tilde{x}|\psi)|P_1)}{2Var(\tilde{x}|\psi)}}} \exp\left\{ - \frac{\frac{Var(E(\tilde{x}|\psi)|P_1)}{2Var(\tilde{x}|\psi)}}{1 + 2\frac{Var(E(\tilde{x}|\psi)|P_1)}{2Var(\tilde{x}|\psi)}} \frac{(E(\tilde{x}|P_1) - P_1)^2}{Var(E(\tilde{x}|\psi)|P_1)} \right\}$$

$$= \sqrt{\frac{Var(\tilde{x}|\psi)}{Var(\tilde{x}|P_1)}} \exp\left\{ - \frac{(E(\tilde{x}|P_1) - P_1)^2}{2Var(\tilde{x}|P_1)} \right\},$$

where the last equality follows from $Var(\tilde{x} \mid \psi) + Var(E(\tilde{x} \mid \psi) \mid P_1) = Var(\tilde{x} \mid P_1)$. With this we can compute the expected utility of an informed trader conditional on P_1:

$$E[u^I(D_1^I)|P_1] = -e^{-\rho W_1^I} E[e^{-(E(\tilde{x}|\psi) - P_1)^2/2Var(\tilde{x}|\psi)}|P_1] =$$

$$-e^{-\rho W_1^I} \sqrt{\frac{Var(\tilde{x}|\psi)}{Var(\tilde{x}|P_1)}} e^{-(E(\tilde{x}|P_1) - P_1)^2/2Var(\tilde{x}|P_1)}.$$

Next, we derive the first-order condition for the informed and uninformed investors' holdings at $t = 0$. The first derivative of the uninformed investor's expected utility with respect to D_0^U is:

$$\frac{\partial E[u^U(D_1^U)]}{\partial D_0^U} = \frac{\partial}{\partial D_0^U} \int_{P_1} E[u^U(D_1^U)|P_1]\phi_{P_1}(P_1)dP_1 = \int_{P_1} \frac{\partial}{\partial D_0^U} E[u^U(D_1^U)|P_1]\phi_{P_1}(P_1)dP_1,$$

where $\phi_{P_1}(\cdot)$ is the pdf of $\tilde{P}_1$. Hence, the FOC for the uninformed investor's holding at $t = 0$ is:

$$e^{-\rho W_0^U} \int_{P_1} \rho(P_1 - P_0) e^{-\rho D_0^U (P_1 - P_0)} e^{-(E(\tilde{x}|P_1) - P_1)^2/2Var(\tilde{x}|P_1)} \phi_{P_1}(P_1)dP_1 = 0.$$

The first derivative of the informed investor's expected utility with respect to D_0^I is:

$$\frac{\partial E[u^I(D_1^I)]}{\partial D_0^I} = \frac{\partial}{\partial D_0^I} \int_{P_1} E[u^I(D_1^I)|P_1]\phi_{P_1}(P_1)dP_1 = \int_{P_1} \frac{\partial}{\partial D_0^I} E[u^I(D_1^I)|P_1]\phi_{P_1}(P_1)dP_1$$

and, hence, the FOC for the informed investor's holding at $t = 0$ is:

$$e^{-\rho W_0^I} \sqrt{\frac{Var(\tilde{x}|\psi)}{Var(\tilde{x}|P_1)}} \int_{P_1} \rho(P_1 - P_0) e^{-\rho D_0^I (P_1 - P_0)} e^{-(E(\tilde{x}|P_1) - P_1)^2/2Var(\tilde{x}|P_1)} \phi_{P_1}(P_1)dP_1 = 0.$$

In addition, for $D_0^i \to -\infty(\infty)$, $i=I,U$, an investor's marginal expected utility with respect to D_0^i converges to ∞ $(-\infty)$ and her expected utility converges to $-\infty$. This, together with the fact that the second derivative of the expected utility is negative for all D_0^i, implies that there exists a unique optimal $D_0^i(P_0)$ for both informed and uninformed investors. Since both FOCs are identical up to a constant multiplier, the optimal demand function $D_0(P_0)$ is the same for both informed and

uninformed investors. In particular, the market-clearing condition in (1) yields:

$$D_0 = \mu_s.$$

Proof of Lemma 2

The proof of Lemma 2 follows along the same lines as the proof of Lemma 3, part (iii) and is omitted for brevity.

Proof of Lemma 3

Part (i)

$s^*(\psi)$ is given in Equation (5). For $s = s^*(\psi)$, no trade occurs—i.e., $D_1^I(\psi, P_1(\psi, s^*(\psi))) = D_0^I$. Both integrands in (6) are zero for $s = s^*(\psi)$ and strictly positive elsewhere for any Ψ. Hence, the expected trading volume is strictly positive for any Ψ.

Part (ii)

The first derivative of the expected trading volume with respect to Ψ is:

$$\begin{aligned}\frac{\partial E[\lambda|D_1^I(\psi,P_1) - D_0^I||\psi]}{\partial\psi} &= -\lambda\int_{-\infty}^{s^*(\psi)}\frac{\partial}{\partial\psi}D_1^I(\psi,P_1)\phi_s(s)ds + \lambda\frac{\partial s^*(\psi)}{\partial\psi}[D_0^I \\ &\quad - D_1^I(\psi,P_1(\psi,s^*(\psi)))]\phi_s(s^*(\psi)) + \lambda\int_{s^*(\psi)}^{\infty}\frac{\partial}{\partial\psi}D_1^I(\psi,P_1)\phi_s(s)ds \\ &\quad - \lambda\frac{\partial s^*(\psi)}{\partial\psi}[D_1^I(\psi,P_1(\psi,s^*(\psi))) - D_0^I]\phi_s(s^*(\psi)) \\ &= -\lambda\int_{-\infty}^{s^*(\psi)}(1-\alpha_1)\tau_\varepsilon\phi_s(s)ds + 0 + \lambda\int_{s^*(\psi)}^{\infty}(1-\alpha_1)\tau_\varepsilon\phi_s(s)ds - 0 \\ &= \lambda(1-\alpha_1)\tau_\varepsilon[1 - 2\Phi_s(s^*(\psi))].\end{aligned}$$

Straightforward algebra shows that $\alpha_1 < 1$. This together with $E[\tilde{x}\mid\psi] = \psi$ implies that $s^*(\psi)$ is decreasing in Ψ. Hence, the first derivative is negative for Ψ lower than some threshold and positive for Ψ higher than the threshold. The second derivative of the expected trading volume with respect to Ψ yields:

$$\frac{\partial^2 E[\lambda|D_1^I(\psi,P_1) - D_0^I||\psi]}{\partial\psi^2} = -2\lambda(1-\alpha_1)\tau_\varepsilon\phi_s(s^*(\psi))\frac{\partial s^*(\psi)}{\partial\psi} > 0$$

which follows again from $s^*(\psi)$ being decreasing in Ψ.

Part (iii)

We prove the symmetry by showing that the derivative of the expected trading volume with respect to Ψ is of the same magnitude but of opposite sign for $\psi^* + v$ and $\psi^* - v$ for all $v > 0$. First, we show that $s^*(\psi^*) = \mu_s$ where $\psi^* = \mu_x - \frac{\rho\mu_s}{\tau_\psi}$:

$$\begin{aligned}s^*(\psi^*) &= \frac{1}{\alpha_2}(\rho Var(\tilde{x}|\psi^*)D_0 - E(\tilde{x}|\psi^*) + \alpha_0 + \alpha_1\psi^*) \\ &= \frac{1}{\alpha_2}\left(\frac{\rho}{\tau_\varepsilon}\mu_s - (1-\alpha_1)\left(\mu_x - \frac{\rho\mu_s}{\tau_\psi}\right) + \alpha_0\right) = \mu_s\end{aligned}$$

where the last equality follows from substituting the expressions for α_0 and α_1 from (13). Then,

$s^*(\psi^* \pm v) = \mu_s \pm (\alpha_1 - 1)v$. For $\psi = \psi^* + v$ the first derivative of the expected trading volume yields:

$$\frac{\partial \lambda E[|D_1^I(\psi, \tilde{P}_1) - D_0^I||\psi = \psi^* + v]}{\partial \psi} = \lambda(1 - \alpha_1)\tau_\varepsilon[1 - 2\Phi_s(\mu_s + (\alpha_1 - 1)v)].$$

For $\psi = \psi^* - v$, the first derivative of the expected trading volume similarly yields:

$$\begin{aligned}\frac{\partial \lambda E[|D_1^I(\psi, \tilde{P}_1) - D_0^I||\psi = \psi^* - v]}{\partial \psi} &= \lambda(1 - \alpha_1)\tau_\varepsilon[1 - 2\Phi_s(\mu_s - (\alpha_1 - 1)v)] \\ &= -\lambda(1 - \alpha_1)\tau_\varepsilon[1 - 2\Phi_s(\mu_s + (\alpha_1 - 1)v)].\end{aligned}$$

Hence, the derivative of the expected trading volume with respect to Ψ is of the same magnitude but of opposite sign for $\psi^* + v$ and $\psi^* - v$. This proves that the expected trading volume is symmetric around ψ^*. As a result, the expected trading volume has its unique local minimum at ψ^*.

Part (iv)

The limit of the slope of the expected trading volume as $\Psi \to \infty$ is:

$$\lim_{\psi \to \infty} \frac{\partial \lambda E[|D_1^I(\psi, \tilde{P}_1) - D_0^I||\psi]}{\partial \psi} = \lambda(1 - \alpha_1)\tau_\varepsilon.$$

From symmetry, it follows that:

$$\lim_{\psi \to -\infty} \frac{\partial \lambda E[|D_1^I(\psi, \tilde{P}_1) - D_0^I||\psi]}{\partial \psi} = -\lambda(1 - \alpha_1)\tau_\varepsilon.$$

Proof of Proposition 1—Analyst's Forecasting Strategy

$$u^A(x^R, \psi) = \lambda c_A E[|D_1^I(x^R, P_1(x^R, \tilde{s})) - D_0^I||x^R] - \frac{1}{2}E[(x^R - \tilde{x})^2|\psi]$$

Let $\hat{\psi}(x^R)$ denote the signal Ψ that informed investors infer from a report x^R:

$$\begin{aligned}E[|D_1^I(\hat{\psi}(x^R), P_1(\hat{\psi}(x^R), \tilde{s})) - D_0^I||x^R] = &\int_{-\infty}^{s^*(\hat{\psi}(x^R))} (D_0^I - D_1^I(\hat{\psi}(x^R), P_1(\hat{\psi}(x^R), \tilde{s})))\phi(s)ds \\ &+ \int_{s^*(\hat{\psi}(x^R))}^{\infty} (D_1^I(\hat{\psi}(x^R), P_1(\hat{\psi}(x^R), \tilde{s})) - D_0^I)\phi(s)ds\end{aligned}$$

The first-order condition yields:

$$\frac{\partial \lambda c_A E[|D_1^I(\hat{\psi}(x^R), P_1(\hat{\psi}(x^R), \tilde{s})) - D_0^I||x^R]}{\partial x^R} - E[(x^R - \tilde{x})|\psi] = 0$$

$$\frac{\partial \lambda c_A E[|D_1^I(\hat{\psi}(x^R), P_1(\hat{\psi}(x^R), \tilde{s})) - D_0^I||x^R]}{\partial \hat{\psi}} \frac{\partial \hat{\psi}(x^R)}{\partial x^R} - E[(x^R - \tilde{x})|\psi] = 0$$

$$\lambda c_A(1-\alpha_1)\tau_\varepsilon[1-2\Phi_s(s^*(\hat{\psi}(x^R)))]\frac{\partial\hat{\psi}(x^R)}{\partial x^R}-E[(x^R-\tilde{x})|\psi]=0.$$

For notational convenience, let $g(\psi)\equiv\lambda c_A(1-\alpha_1)\tau_\varepsilon[1-2\Phi_s(s^*(\psi))]$. In equilibrium, we have $\hat{\psi}(x^R(\psi))=\psi$ and, hence, $\hat{\psi}'(x^R)=1/\partial x^R(\psi)/\partial\psi$. Moreover, $\partial x^R(\psi)/\partial\psi=b'(\psi)+1$ and $E[(x^R-\tilde{x})|\psi]=b(\psi)$. So, the first-order condition yields:

$$g(\psi)-(b'(\psi)+1)b(\psi)=0 \text{ with } b(\psi^*)=0. \quad (14)$$

The first-order condition provides an ordinary differential equation (ODE) for the analyst's bias function $b(\Psi)$. Claim 1 provides the boundary condition $b(\psi^*)=0$. Next, we substitute $c(\psi)=1/2(b(\psi))^2$. We begin by considering the case of $\psi\geq\psi^*$. From Claim 3, we know that $b\geq 0$. Then, $b(\psi)=\sqrt{2|c(\psi)|}$ and the first-order condition in (14) can be rewritten as:

$$c'(\psi)=g(\psi)-\sqrt{2|c(\psi)|} \text{ with } c(\psi^*)=0. \quad (15)$$

Since the ODE is not Lipschitz continuous at ψ^*, we appeal to the theorem by Cauchy-Peano for existence of a solution.[22] The Cauchy-Peano Theorem requires that the right-hand side of the ODE is continuous. This is the case for the ODE in (15) and, hence, there exists a function $c(\Psi)$ satisfying the initial value problem in (15).

Next, we show that $c(\Psi)$ provides a solution to the analyst's forecasting problem. That requires $c(\Psi)\geq 0$ such that $b(\Psi)$ is real. Note that if $c(\Psi)=0$ for some $\psi>\psi^*$, then $c'(\psi)>0$ because $g(\Psi)>0$ for all $\psi>\psi^*$ (recall that $s^*(\psi^*)=\mu_s$ and $s^*(\psi)$ is decreasing in Ψ). This means that if $c(\psi_1)=0$ for some $\psi_1>\psi^*$, then $c(\Psi)\geq 0$ for all $\psi>\psi_1$. Since $c(\psi^*)=0$, we know $c(\Psi)\geq 0$ for all $\psi\geq\psi^*$. We next want to show that $c(\Psi)>0$ for $\psi>\psi^*$. Suppose that $c(\psi_1)=0$ for $\psi_1>\psi^*$. From $g(\psi_1)>0$, it follows that $c'(\psi_1)>0$, which implies that $c(\Psi)<0$ for some $\psi<\psi_1$, a contradiction. The solution to the initial value problem in (15) is therefore strictly positive for $\psi>\psi^*$. A similar argument (using $g(\psi)<0$ for $\psi<\psi^*$) shows that the solution c is positive for $\psi<\psi^*$ as well.

Since $c(\Psi)\geq 0$, we may now write $b(\Psi)$ as $b(\psi)=\sqrt{2c(\psi)}$ for $\psi>\psi^*$ and $b(\psi)=-\sqrt{2c(\psi)}$ for $\psi<\psi^*$ (see Claim 3). As a final step, we show that the derivative of $c(\Psi)$ is continuous. We again start by focusing on $\psi\geq\psi^*$. Since $b'(\psi)=\frac{c'(\psi)}{\sqrt{2c(\psi)}}$ and $c(\psi)>0$ for $\psi>\psi^*$, $b'(\psi)$ is continuous for $\psi>\psi^*$. Suppose that $b(\Psi)$ is not bounded as $\psi\rightarrow\psi^*$. This means that there is a sequence $\psi_k\downarrow\psi^*$ such that $b'(\psi_k)\rightarrow\infty$, and so:

$$\frac{b(\psi_k)}{\psi_k-\psi^*}\rightarrow\infty \text{ as } k\rightarrow\infty.$$

Dividing Equation (14) by $(\psi_k-\psi^*)$ and rearranging yields:

$$\frac{g(\psi_k)}{\psi_k-\psi^*}=(b'(\psi_k)+1)\frac{b(\psi_k)}{\psi_k-\psi^*}.$$

Taking limits on both sides yields $g'(\psi^*)=\infty$, which is a contradiction because we know that:

---

[22] Cauchy-Peano Existence Theorem: Suppose that F is a function defined on $I_t\times U_p$, where I is an open interval in $\mathbb{R}$ and U is an open set in $\mathbb{R}^n$. Suppose now that F is continuously differentiable in t and continuous (not necessarily Lipschitz) in p. Then given any $t_0\in I$ and $p\in U$, there exists an $\varepsilon>0$ such that the initial value problem, $x'(t)=F(t,x(t))$, $x(t_0)=p$, has a solution defined on the interval $(t_0-\varepsilon, t_0+\varepsilon)$. For a proof of the theorem, see Coddington and Levinson (1955, 6).

$$g'(\psi^*) = 2\lambda(1-\alpha_1)\tau_\varepsilon \phi(0)\frac{1}{\sigma_s}\frac{1-\alpha_1}{\alpha_2}.$$

From this it follows that there exists a constant C such that $b'(\psi)$ is bounded by C and $b(\Psi)$ is bounded by $C\ (\psi - \psi^*)$ as $\psi_k \downarrow \psi^*$. We next show that $b'(\psi)$ has a limit. Taking limits, we can see that $g'(\psi^*) = \beta(1+\beta)$, where $\beta \equiv \lim_{\psi_k \downarrow \psi^*} \frac{b(\psi_k)}{\psi_k - \psi^*}$. This equation has two solutions, only one of which is positive (we know from Claim 4 that $b'(\psi) \geq 0$). To conclude that $b'(\psi)$ is continuous at ψ^*, we now observe that for any sequence $\psi_k \downarrow \psi^*$, we may extract a convergent subsequence and this subsequence must converge to the solution for β. Thus $b'(\psi) \rightarrow \beta$ as $\psi_k \downarrow \psi^*$ and so $b'(\psi)$ is continuous from the right at ψ^*.

An identical argument shows that b is defined for $\psi \leq \psi^*$ and continuous from the left at ψ^*. The value for $b'(\psi^*)$ will again be the positive root of the quadratic equation $g'(\psi^*) = \beta(1 + \beta)$, such that $b'(\psi)$ will also be continuous. This completes the proof.

Empirical Predictions

Proof of Corollary 1

Since the expected trade function is symmetric with respect to ψ^* and the expected forecast costs are symmetric around 0, the bias function $b(\Psi)$ is symmetric around ψ^* and $b\ (\psi) = -b\ (2\psi^* - \psi)$. Given that $\psi^* < \mu_x$ and the symmetry of the distribution function of Ψ around μ_x the pdf of Ψ is higher than the pdf of $2\psi^* - \psi$ for any $\psi > \psi^*$. Hence, the expected bias is positive.

Proof of Corollary 2

A sufficient condition for the corollary to hold is that the analyst's bias is positive for $\psi > \psi^*$ and negative for $\psi < \psi^*$. Proposition 1 states that this condition holds in equilibrium, which completes the proof of the corollary.

Proof of Corollary 3

Part (a). Since the marginal benefit from increasing the forecast is linear in c_A and the marginal cost is independent of c_A, the magnitude of the analyst's bias, $|b(\psi)|$, is increasing in c_A for all $\psi \neq \psi^*$.

Part (b). We first want to show that $\frac{\partial E[-FE]}{\partial c_A} = \frac{\partial E[b(\psi)]}{\partial c_A} > 0$. We have:

$$E[b(\tilde{\psi})] = \int_{\psi^*}^{\infty} b(\psi; c_A)[f(\psi) - f(2\psi^* - \psi)]d\psi$$

and, hence:

$$\frac{\partial E[b(\tilde{\psi})]}{\partial c_A} = \int_{\psi^*}^{\infty} \frac{\partial b(\psi; c_A)}{\partial c_A}[f(\psi) - f(2\psi^* - \psi)]d\psi > 0$$

because $\frac{\partial b(\psi, c_A)}{\partial c_A} > 0$ and both $[f\ (\psi) - f\ (2\psi^* - \psi)]$ and ψ^* are independent of c_A.

Next, we show that $\partial E[FE^2]/\partial c_A > 0$. Since the magnitude of $b(\Psi)$ is increasing in c_A for all $\psi \neq \psi^*$, the expected squared bias is also increasing in c_A. The conditional variance of x does not depend on c_A.

Part (c). Since the marginal benefit from increasing the forecast is finite, the analyst's marginal benefit approaches 0 as $c_A \rightarrow 0$. It follows that, in equilibrium, the analyst's expected marginal cost from biasing the forecast also approaches zero as $c_A \rightarrow 0$ and, hence, $\lim_{c_A \rightarrow 0} b(\psi) = 0$.

Proof of Corollary 4

As claimed before, as the precision of the private information, τ_ε, goes to zero, the expected bias of the analyst goes to zero as well (since the trade generated by the analyst's forecast goes to zero). This holds for any finite weight that the analyst's utility assigns to the trading volume, c_A. Formally, $\lim_{\tau_\varepsilon \to 0} b(\psi) = 0$ for any $c_A \in [0, \infty)$ and any Ψ. Also, $\lim_{c_A \to \infty} E[b(\tilde{\psi})] = \infty$ for any $\tau_\varepsilon > 0$ and, hence, $\lim_{c_A \to \infty} E[b(\tilde{\psi})^2] = \infty$. Moreover, for any $\psi \neq \psi^*$ we know from Corollary 3 that $b(\psi)^2$ is increasing in c_A. From continuity it follows that there exists a sufficiently high c_A^*, such that for all $c_A > c_A^*$ the expected squared bias increases in τ_ε, for sufficiently small values of τ_ε, faster than the decrease in $Var(\tilde{x} \mid \psi) = 1/\tau_\varepsilon$. This, together with Equation (7) implies that there exists c_A^* such that for all $c_A > c_A^*$ the expected squared forecast error increases in τ_ε for sufficiently small values of τ_ε.

Proof of Proposition 2

Informed Investor's Demand Function at t = 0

We can compute $D_1^I(\psi)$ as:

$$D_1^I(\psi) = \begin{cases} \dfrac{E[\tilde{x}|x^R] - P_0 - c_I}{\rho/\tau_\varepsilon} & \text{if } \psi > \psi^* + c_I \\ D_0^I & \text{if } \psi^* + c_I > \psi > \psi^* - c_I \\ \dfrac{E[\tilde{x}|x^R] - P_0 + c_I}{\rho/\tau_\varepsilon} & \text{if } \psi^* - c_I > \psi, \end{cases}$$

where:

$$\psi^* = \frac{D_0 \rho}{\tau_\varepsilon} + P_0. \tag{16}$$

For notational convenience, let the no-trading zone be denoted by $\Psi_{NT} = [\underline{\psi}, \bar{\psi}]$, where $\underline{\psi} = \psi^* - c_I$ and $\bar{\psi} = \psi^* + c_I$. Given the optimal $D_1^I(\psi)$, we can compute the informed investor's *ex ante* expected utility as:

$$-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\rho[W_0 + (\tilde{x} - P_0)D_1^I - c_I|D_1^I - D_0^I|]} f(x|\psi)h(\psi)dx d\psi =$$

$$-\int_{-\infty}^{\underline{\psi}} e^{-\rho\left[W_0 + (E[\tilde{x}|\psi] - P_0)\frac{E[\tilde{x}|\psi] - P_0 + c_I}{\rho Var(\tilde{x}|\psi)} + c_I\left(\frac{E[\tilde{x}|\psi] - P_0 + c_I}{\rho Var(\tilde{x}|\psi)} - D_0^I\right) - \frac{\rho}{2}\left(\frac{E[\tilde{x}|\psi] - P_0 + c_I}{\rho Var(\tilde{x}|\psi)}\right)^2 Var(\tilde{x}|\psi)\right]} h(\psi)d\psi$$

$$-\int_{\underline{\psi}}^{\bar{\psi}} e^{-\rho\left[W_0 + (E[\tilde{x}|\psi] - P_0)D_0^I - \frac{\rho}{2}D_0^2 Var(\tilde{x}|\psi)\right]} h(\psi)d\psi$$

$$-\int_{\bar{\psi}}^{\infty} e^{-\rho\left[W_0 + (E[\tilde{x}|\psi] - P_0)\frac{E[\tilde{x}|\psi] - P_0 + c_I}{\rho Var(\tilde{x}|\psi)} - c_I\left(\frac{E[\tilde{x}|\psi] - P_0 + c_I}{\rho Var(\tilde{x}|\psi)} - D_0^I\right) - \frac{\rho}{2}\left(\frac{E[\tilde{x}|\psi] - P_0 + c_I}{\rho Var(\tilde{x}|\psi)}\right)^2 Var(\tilde{x}|\psi)\right]} h(\psi)d\psi, \tag{17}$$

where $f(x|\psi)$ is the conditional distribution of x and $h(\Psi)$ is the prior distribution of Ψ. Let:

$$MV(\psi)|_{c_I=0} = W_0 + \frac{1}{2}\frac{(E[\tilde{x}|\psi] - P_0)^2}{\rho Var(\tilde{x}|\psi)}$$

denote the mean-variance term if $c_I = 0$. WLOG, we set $W_0 = 0$ for the remainder of the proof. We can then rewrite (17) as:

$$-\int_{-\infty}^{\underline{\psi}} e^{-\rho[MV(\psi)|_{c_I=0}-c_I(D_0^I-E[\tilde{x}|\psi]-P_0+1/2c_I/\rho Var(\tilde{x}|\psi))]}h(\psi)d\psi$$

$$-\int_{\underline{\psi}}^{\bar{\psi}} e^{-\rho[MV(\psi)|_{c_I=0}+(E[\tilde{x}|\psi]-P_0-\rho/2D_0^I Var(\tilde{x}|\psi))D_0^I-1/2(E[\tilde{x}|\psi]-P_0)^2/\rho Var(\tilde{x}|\psi)]}h(\psi)d\psi$$

$$-\int_{\bar{\psi}}^{\infty} e^{-\rho[MV(\psi)|_{c_I=0}-c_I(E[\tilde{x}|\psi]-P_0-1/2c_I\rho Var(\tilde{x}|\psi)-D_0^I)]}h(\psi)d\psi. \tag{18}$$

The informed investor chooses D_0^I to maximize her *ex ante* expected utility, taking into account her optimal trading strategy at $t = 1$. Equation (16) shows a one-to-one relation between ψ^* and D_0^I. In particular, $D_0^I = E[\tilde{x} \mid \psi^*] - P_0 / \rho Var(\tilde{x} \mid \psi)$. Hence, solving for the optimal D_0 is equivalent to solving for the optimal ψ^*. Rewriting (18) yields:

$$-\int_{-\infty}^{\underline{\psi}} e^{-\rho[MV(\psi)|_{c_I=0}-c_I(E[\tilde{x}|\psi^*]-E[\tilde{x}|\psi]-1/2c_I/\rho Var(\tilde{x}|\psi))]}h(\psi)d\psi$$

$$-\int_{\underline{\psi}}^{\bar{\psi}} e^{-\rho[MV(\psi)|_{c_I=0}-1/2(E[\tilde{x}|\psi^*]-E[\tilde{x}|\psi])^2/\rho Var(\tilde{x}|\psi)]}h(\psi)d\psi$$

$$-\int_{\bar{\psi}}^{\infty} e^{-\rho[MV(\psi)|_{c_I=0}-c_I(E[\tilde{x}|\psi]-E[\tilde{x}|\psi^*]-1/2c_I/\rho Var(\tilde{x}\,\psi))]}h(\psi)d\psi.$$

Further:

$$\begin{aligned}
e^{-\rho MV(\psi)|_{c_I=0}}h(\psi) &\propto \exp\left\{-\frac{1}{2}\frac{(E[\tilde{x}|\psi]-P_0)^2}{Var(\tilde{x}|\psi)}\right\}\exp\left\{-\frac{\tau_\psi}{2}(\psi-\mu_x)^2\right\} \\
&= \exp\left\{-\frac{1}{2}(\tau_\varepsilon(\psi-P_0)^2+\tau_\psi(\psi-\mu_x)^2)\right\} \\
&= \exp\left\{-\frac{1}{2}((\tau_\varepsilon+\tau_\psi)\psi^2+\tau_\psi\mu_x^2+\tau_\varepsilon P_0^2-2P_0\psi\tau_\varepsilon-2\mu_x\psi\tau_\psi)\right\} \\
&= \exp\left\{-\frac{1}{2}\left((\tau_\varepsilon+\tau_\psi)\left(\psi-\frac{P_0\tau_\varepsilon+\mu_x\tau_\psi}{\tau_\varepsilon+\tau_\psi}\right)^2+\frac{\tau_\varepsilon\tau_\psi}{\tau_\varepsilon+\tau_\psi}(\mu_x-P_0)^2\right)\right\} \\
&= \exp\left\{-\frac{1}{2}\left((\tau_\varepsilon+\tau_\psi)\left(\psi-\frac{P_0\tau_\varepsilon+\mu_x\tau_\psi}{\tau_\varepsilon+\tau_\psi}\right)^2\right)\right\}\exp\left\{-\frac{1}{2}\frac{\tau_\varepsilon\tau_\psi}{\tau_\varepsilon+\tau_\psi}(\mu_x-P_0)^2\right\} \\
&\propto \phi\left(\sqrt{\tau_\varepsilon+\tau_\psi}\left(\psi-\frac{P_0\tau_\varepsilon+\mu_x\tau_\psi}{\tau_\varepsilon+\tau_\psi}\right)\right),
\end{aligned}$$

where $\phi()$ denotes the pdf of a standard normal distribution. Hence, the informed investor's objective is to choose ψ^* that maximizes:

$$\max_{\psi^*}\int_{-\infty}^{\infty}\delta(\psi,\psi^*)\phi\left(\sqrt{\tau_\varepsilon+\tau_\psi}\left(\psi-\frac{P_0\tau_\varepsilon+\mu_x\tau_\psi}{\tau_\varepsilon+\tau_\psi}\right)\right)d\psi, \tag{19}$$

where $\delta(\psi, \psi^*)$ is continuous and symmetric with respect to ψ^* with a unique global maximum at $\psi = \psi^*$:

$$\delta(\psi,\psi^*) = \begin{cases} -e^{c^I E[\tilde{x}|\psi^*] - E[\tilde{x}|\psi] - 1/2c_I/Var(\tilde{x}|\psi)} & \text{if } \psi < \underline{\psi} \\ -e^{(E[\tilde{x}|\psi] - E[\tilde{x}|\psi^*])^2/2Var(\tilde{x}|\psi)} & \text{if } \underline{\psi} \leq \psi \leq \overline{\psi} \\ -e^{c^I E[\tilde{x}|\psi] - E[\tilde{x}|\psi^*] - 1/2c_I/Var(\tilde{x}|\psi)} & \text{if } \overline{\psi} < \psi \end{cases}$$

It follows that the value of ψ^* that maximizes (19) equals $\frac{P_0\tau_\varepsilon+\mu_x\tau_\psi}{\tau_\varepsilon+\tau_\psi}$. Solving (16) for $\psi^* = \frac{P_0\tau_\varepsilon+\mu_x\tau_\psi}{\tau_\varepsilon+\tau_\psi}$ yields $D_0^I = D_0^U = \mu_s$.

Analyst's Forecasting Strategy

We start by solving for the equilibrium-forecasting strategy and the investor's trade for the case where $\psi > \psi^* + c_I$ and showing its uniqueness. For $\psi > \psi^* + c_I$, the analyst maximizes:

$$c_A \frac{E[\tilde{x}|\hat{\psi}(x^R)] - P_0 - c_I}{\rho Var(\tilde{x}|\psi)} - \frac{1}{2}E[(x^R - \tilde{x})^2|\psi],$$

where $\hat{\psi}(x^R)$ denotes the informed investor's beliefs about the analyst's private signal Ψ following the forecast x^R. The FOC is:

$$0 = c_A \frac{\hat{\psi}'(x^R)}{\rho Var(\tilde{x}|\psi)} - E[x^R - \tilde{x}|\psi]$$

$$\Leftrightarrow 0 = c_A \frac{\hat{\psi}'(x^R)}{\rho Var(\tilde{x}|\psi)} - b(\psi). \tag{20}$$

In equilibrium, the informed investor's beliefs about the analyst's forecasting strategy has to be correct—i.e., $\hat{\psi}(x^R(\psi)) = \psi$. Then, $\hat{\psi}'(x^R) = 1/\partial x^R(\psi)/\partial\psi = 1/1 + b'(\psi)$. With $Var(\tilde{x}\mid\psi) = 1/\tau_\varepsilon$, we can rewrite (20) as:

$$0 = \frac{c_A}{\rho}\frac{\tau_\varepsilon}{1 + b'(\psi)} - b(\psi)$$

$$\Leftrightarrow 0 = b(\psi) + b'(\psi)b(\psi) - \frac{c_A\tau_\varepsilon}{\rho}.$$

Hence, the FOC is a differential equation for $b(\Psi)$. With the boundary condition that $b(\psi^* + c_I) = 0$, the Fundamental Theorem of Differential Equations guarantees that the solution exists and is unique. The unique solution is:

$$b(\psi) = \frac{c_A\tau_\varepsilon}{\rho}(1 + PL(-e^{-(\psi-c_I-\psi^*)\rho/c_A^2\tau_\varepsilon-1})) \text{ for } \psi > \psi^* + c_I, \tag{21}$$

where $PL()$ denotes the *ProductLog* function (see below).

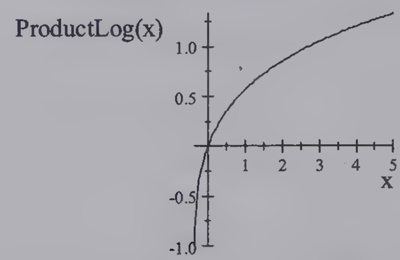
For $\psi < \psi^* - c_I$ the differential equation for $b(\Psi)$ is:

$$0 = b(\psi) + b'(\psi)b(\psi) + \frac{c_A\tau_\varepsilon}{\rho}.$$

The boundary condition for $\psi < \psi^* - c_I$ is $b(\psi^* - c_I) = 0$ and, hence:

$$b(\psi) = -\frac{c_A\tau_\varepsilon}{\rho}(1 + PL(-e^{-(\psi^*-\psi)\rho/c_A\tau_\varepsilon-1})) \text{ for } \psi < \psi^* - c_I.$$

FIGURE 4
ProductLog(x)



ProductLog(x)

The $ProductLog(x)$ function (also known as Lambert-W or Omega function) is the solution to the differential equation $f'(x) = \frac{f(x)}{x(1+f(x))}$. It is defined as any $x \geq -e^{-1}$, where $ProductLog(-e^{-1}) = -1$, $ProductLog(0) = 0$, and $ProductLog(x) = \infty$. Figure 4 demonstrates the $ProductLog(x)$ function.

REFERENCES

Bernhardt, D., M. Campello, and E. Kutsoati. 2006. Who herds? *Journal of Financial Economics* 80 (3): 657–675.

Beyer, A. 2008. Financial analysts' forecast revisions and managers' reporting behavior. *Journal of Accounting and Economics* 46 (2-3): 334–348.

Chen, Q., and W. Jiang. 2005. Analysts' weighting of private and public information. *Review of Financial Studies* 19 (1): 319–355.

Coddington, E. A., and N. Levinson. 1955. *Theory of Ordinary Differential Equations*. New York, NY: McGraw-Hill.

Cowen, A., B. Groysberg, and P. Healy. 2006. Which types of analyst firms are more optimistic? *Journal of Accounting and Economics* 41 (1-2): 119–146.

Dugar, A., and S. Nathan. 1995. The effect of investment banking relationships on financial analysts' earnings forecasts and investment recommendations. *Contemporary Accounting Research* 12 (1): 131–160.

Dye, R. A., and S. Sridhar. 2004. Reliability-relevance trade-offs and the efficiency of aggregation. *Journal of Accounting Research* 42 (1): 51–88.

Easterwood, J., and S. Nutt. 1999. Inefficiency in analysts' earnings forecasts: Systematic misreaction or systematic optimism? *The Journal of Finance* 54 (5): 1777–1797.

Fischer, P., and P. Stocken. 2010. Analyst information acquisition and communication. *The Accounting Review* 85 (6): 1985–2009.

———, and R. E. Verrecchia. 2000. Reporting bias. *The Accounting Review* 75 (2): 229–245.

Friesen, G., and P. Weller. 2006. Quantifying cognitive biases in analyst earnings forecasts. *Journal of Financial Markets* 9 (4): 333–365.

Grossman, S. J., and J. E. Stiglitz. 1980. On the impossibility of informationally efficient markets. *The American Economic Review* 70 (3): 393–408.

Gu, Z., and J. Xue. 2008. The superiority and disciplining role of independent analysts. *Journal of Accounting and Economics* 45 (2-3): 289–316.

Guttman, I. 2010. The timing of analysts' earnings forecasts. *The Accounting Review* 85 (2): 513–545.

Hayes, R. 1998. The impact of trading commission incentives on analyst's stock coverage decision and earnings forecasts. *Journal of Accounting Research* 36 (2): 299–320.

Hong, H., and J. Kubik. 2003. Analyzing the analysts: Career concerns and biased earnings forecasts. *The Journal of Finance* 58 (1): 313–351.

Irvine, P. 2004. Analysts' forecasts and brokerage-firm trading. *The Accounting Review* 79 (1): 125–149.

Jackson, A. 2005. Trade generation, reputation, and sell-side analysts. *The Journal of Finance* 60 (2): 673–717.

Kothari, S. P. 2001. Capitals research in accounting. *Journal of Accounting and Economics* 31 (1-3): 105–231.

Lim, T. 2001. Rationality and analysts forecast bias. *The Journal of Finance* 56 (1): 369–385.

Lin, H., and M. McNichols. 1998. Underwriting relationships, analysts' earnings forecasts and investment recommendations. *Journal of Accounting and Economics* 25 (1): 101–127.

Michaely, R., and K. Womack. 1999. Conflict of interest and credibility of underwriter analyst recommendations. *Review of Financial Studies* 12 (4): 653–686.

Mikhail, M., B. Walther, and R. Willis. 1999. Does forecast accuracy matter to security analysts? *The Accounting Review* 74 (2): 185–200.

Morgan, J., and P. C. Stocken. 2003. An analysis of stock recommendations. *The Rand Journal of Economics* 34 (1): 183–203.

Stickel, S. E. 1992. Reputation and performance among security analysts. *The Journal of Finance* 47 (5): 1811–1836.

Stocken, P. C. 2000. Credibility of voluntary disclosure. *The Rand Journal of Economics* 31 (2): 359–374.

THE ACCOUNTING REVIEW
Vol. 86, No. 2
2011
pp. 483–505

American Accounting Association
DOI: 10.2308/accr.00000018

Earnings Quality and International IPO Underpricing

Thomas J. Boulton
Miami University

Scott B. Smart
Indiana University

Chad J. Zutter
University of Pittsburgh

ABSTRACT: This study examines the impact of country-level earnings quality on IPO underpricing. Examining 10,783 IPOs from 37 countries, we find that IPOs are underpriced less in countries where public firms produce higher quality earnings information. This finding persists after controlling for other deal- and country-specific factors that affect IPO underpricing, and it is driven neither by the large and relatively transparent markets in the U.S. and U.K. nor by the relatively opaque Japanese market. The impact of low earnings quality on underpricing is partially offset by the use of a top-tier underwriter.

Keywords: *international finance; earnings quality; initial public offerings; underpricing.*

Data Availability: *Data used in this study are available from public sources.*

JEL Classifications: *G15; G24; G30; G32; G34.*

I. INTRODUCTION

Few corporate events garner more attention from researchers, practitioners, the media, and the public than initial public offerings (IPOs). Generally, the focus is on the large, sometimes spectacular, first-day gains to new issues observed not only in the U.S., but also in virtually all of the world's stock markets. Summarizing the findings of dozens of studies, the

For valuable comments, the authors express their thanks to Steven Kachelmeier (senior editor), Mark Trombley (editor), two anonymous reviewers, Oya Altinkiliç, Utpal Bhattacharya, Ting Chen, Robin Grieves, Robert Jennings, Kenneth Lehn, Gershon Mandelker, Jay Ritter, Shawn Thomas, Greg Udell, Xiaoyun Yu, and seminar participants at the 20th Annual Conference on Financial Economics and Accounting (Rutgers), Financial Management Association meetings (Orlando), Central Michigan University, Duquesne University, Louisiana Tech University, Marquette University, Miami University, Oklahoma State University, Seton Hall University, Texas A&M University, University of Nebraska, The University of Oklahoma, University of Otago, University of Pittsburgh, West Virginia University, and Wright State University. Any remaining errors or omissions remain the responsibility of the authors.

Editor's note: Accepted by Mark Trombley.

Submitted: October 2008
Accepted: August 2010
Published Online: March 2011

majority of which focus on underpricing in a single country, Loughran et al. (1994) confirm that IPOs earn positive first-day returns everywhere, and that underpricing varies dramatically across countries. However, what drives underpricing differences across markets is still a largely unexplored question.

Theoretical explanations of the underpricing phenomenon often propose that underpricing arises from information asymmetries among participants in the IPO process. To cite just one prominent example, Rock (1986) shows that when investors have different information sets, underpricing is necessary to induce less informed investors to bid for IPO shares in equilibrium.[1] Asymmetric information theories receive considerable empirical support. In a recent survey of the empirical evidence, Ljungqvist (2007, 380) concludes that information asymmetries "have a first-order effect on underpricing." We investigate whether some of the cross-country underpricing variation documented in the literature may be driven by differences in the quality of information available to investors in different markets. Specifically, we are interested in whether international differences in the quality of accounting information (i.e., earnings quality) help explain the international cross-section of IPO initial returns (i.e., IPO underpricing).

We draw upon the extant literature for several proxies for earnings quality and to determine whether those proxies influence underpricing across firms in different countries. There is ample evidence in both the accounting and finance literatures that the earnings quality measures in our tests are correlated with a wide range of capital market outcomes.

For example, Bhattacharya et al. (2003) find that, in countries with greater earnings opacity, firms face a higher cost of capital. Biddle and Hilary (2006) establish that better accounting information leads to more efficient firm-level investment decisions, particularly in countries where financing comes through arm's-length transactions rather than through relationships with creditors. Gelos and Wei (2005) find that emerging markets mutual fund managers invest more heavily in countries with greater transparency. Jin and Myers (2006) argue that greater opacity leads to lower firm-specific risk borne by investors and more synchronous stock prices, indicating that the relative amount of firm-level information (as opposed to market-level information) that is capitalized into stock prices is lower in more opaque countries.

These findings, combined with the central role that information asymmetries play in IPO underpricing, give rise to our earnings quality hypothesis:

Hypothesis: Do firms endure higher underpricing when they go public in countries where the quality of earnings information is relatively low?

We approach this question in an international setting for several reasons. First, as noted above, underpricing varies dramatically across countries, though relatively little work has been done to understand why this is the case. Second, within a single country, firms operate in the same legal and regulatory environment, and that environment likely imposes some limit on the variability of earnings quality across firms within a country. Thus, it is plausible that the quality of accounting information varies more across markets than within a single market. Third, in most cases, very little accounting information is available about a particular firm prior to its IPO. An IPO prospectus typically offers little more than two years of pre-IPO financial information. This limits the ability to capture cross-sectional earnings quality differences as of the IPO date. In contrast, the

[1] Other information-based underpricing theories endow different participants in the IPO process with superior information. For example, in Baron (1982) investment bankers know more about an IPO firm's true value than does the issuer, and underpricing becomes a solution to the resulting principal-agent problem. In Allen and Faulhaber (1989), Grinblatt and Hwang (1989), and Welch (1989) issuers use their information advantage to signal firm quality, whereas in Benveniste and Spindt (1989) underpricing induces well-informed investors to reveal what they know before the offer price is set.

country-level earnings quality measures that we use draw upon several years of data generated by many different firms. While these measures do not provide specific information on the quality of earnings from one IPO to another in a given country, they do capture differences in the information environment in which firms from different countries go public. Fourth, we are interested in whether reputable financial intermediaries, such as top-tier investment banks, play a role in mitigating the information asymmetries related to earnings quality. Given the global scope of the investment banking industry, and given the difficulties in identifying cross-sectional earnings quality variance in a single country, the question of whether investment banks mitigate earnings-related information asymmetries seems ideally suited to an international sample.

Although the advantages of studying underpricing using an international sample are numerous, we must mention two caveats associated with our research design. First, underpricing differences across countries could be influenced by omitted variables, such as differences in offering methods, which could be correlated with our earnings quality measures. We have taken many steps to minimize this possibility, for example by controlling for country and deal characteristics such as the offering method. In addition, we appeal to evidence that IPO pricing methods around the world are converging over time, with the U.S. bookbuilding approach dominating in most countries. For instance, Ljungqvist et al. (2003) find that, in a sample of 2,143 IPOs from 65 countries, 46.2 percent were priced using the bookbuilding method in 1994, but by the end of their sample period (the first seven months of 1999), the fraction of new issues priced via bookbuilding rose to 80 percent. Similarly, Jagannathan and Sherman (2006) examine the use of auctions and bookbuilding in 46 countries and find that in all but four countries, auctions have been abandoned entirely and are rare in the few countries that still use them. In contrast, they find that the bookbuilding method has been gaining market share over time and has become the dominant pricing method in most countries. While we cannot rule out the possibility of an omitted variables problem in our analysis, the timeliness of our sample (1998–2008) combined with our regression control variables and robustness tests reduce this problem to the extent possible.

The second caveat is that private firms can presumably mitigate asymmetric information problems by providing high-quality accounting information leading up to their IPOs, but we do not have data on pre-IPO disclosures of private firms. Our country-level measures are intended to quantify, in a broad sense, the quality of information generally provided by public firms in a given country, but these measures may or may not represent a good characterization of the reporting practices of private firms prior to their IPOs. Ball and Shivakumar (2005), for example, find that the reporting done by private U.K. firms more closely resembles the disclosures provided by firms in code-law countries than those of U.K. public companies. Even so, it seems likely that as firms execute their plans to go public, they adapt their reporting practices to align with those of existing public firms. In that case, IPO firms would not resemble other private firms in terms of their reporting strategies. In either case, there may be limits to the benefits that a given firm can capture by adopting reporting practices superior to those that are common in its home country. For example, Doidge et al. (2007) find that firms in countries where corporate governance practices are relatively poor find it difficult to credibly commit themselves to better governance.

Examining 10,783 IPO events, we find evidence of a statistically and economically significant association between country-level earnings quality and IPO-firm-level underpricing. Comparing two countries that differ in earnings quality by one standard deviation reveals that firms going public in the country with better earnings information experience 7.8 percentage points lower underpricing (the sample mean initial return is 36.5 percent).[2] Our regression models control for

[2] This statement is based on the standard deviation of aggregate EM reported in Table 3 and the regression estimate of aggregate EM reported in Table 5.

many deal-specific (e.g., offer size, industry, underwriter, and underwriting method) and country-specific (e.g., market returns, liquidity, and IPO activity) variables that influence underpricing, and our results are robust to accounting for price-stabilization activities and various minimum offer price screens. Furthermore, the link between earnings quality and underpricing is driven neither by the large and relatively transparent U.S. and U.K. markets nor by the relatively opaque market in Japan. Finally, we find evidence that top-tier underwriters mitigate the effect of poor country-level earnings quality on IPO underpricing, such that firms listing in countries with relatively low earnings quality experience a marginal reduction in underpricing if they are backed by top-tier underwriters.

Section II highlights previous research on earnings quality and IPOs related to this study. Section III describes our sample construction and descriptive statistics. Section IV contains our primary results on the relations between the quality of earnings and IPO underpricing and illustrates that many of the conventional variables used in single-country studies of IPO underpricing also explain the international cross-section. Section V summarizes and concludes.

II. EARNINGS QUALITY AND IPOS

A number of researchers have studied the link between accounting information disclosed in the IPO prospectus and the market value of going-public firms. Essentially, this strain of the literature asks whether IPO firms manipulate their financial statements to obtain a higher share price. Early studies offer modest affirmation for this hypothesis (e.g., Clarkson et al. 1992; Aharony et al. 1993; Friedlan 1994). Teoh et al. (1998b) report that IPO firms have high issue-year earnings and abnormal accruals, followed by poor long-run earnings and negative abnormal accruals and that abnormal accruals at the IPO help explain subsequent poor stock returns. Similarly, Teoh et al. (1998a) find that IPO firms that are the most aggressive in using accruals to report cash flows in excess of earnings earn 20 percent lower stock returns in the three years after the IPO compared to the firms reporting the most conservative earnings figures.[3] However, Fan (2007) reports that discretionary accruals are indeed highest in the IPO year, consistent with the notion that IPO issuers manage earnings, but she finds no evidence of a negative relation between IPO earnings management and subsequent stock returns. Further, Venkataraman et al. (2008) examine pre-IPO financial statements and find that pre-IPO accruals tend to be negative and less than post-IPO accruals. In the same vein, Ball and Shivakumar (2008) find that firms' reporting practices become more conservative when they transition from private to public status through an IPO.

All of these studies focus on the extent to which earnings manipulation leading up to the IPO influence the firm's stock price. They do not look at underpricing (i.e., the difference between the IPO offer price and the market price established once trading begins), which is our focus. The accounting literature offers several insights regarding the influence of accounting disclosures on IPO underpricing.

In a study of micro-cap Nasdaq IPOs, Willenborg and McKeown (2001) find that firms going public with going-concern audit opinions are more likely to delist within two years of the IPO, but these firms also endure less underpricing, consistent with the notion that the audit opinion reduces information asymmetries between issuers and investors. Jog and McConomy (2003) examine the impact of voluntary disclosure of management earnings forecasts in the IPO prospectus on IPO valuation and performance. They find higher underpricing for IPOs that do not include earnings forecasts, though this difference is concentrated among small firms. Schrand and Verrecchia (2005) study the relation between underpricing and the frequency of pre-IPO disclosure, finding

[3] Aharony et al. (2000), DuCharme et al. (2001), and Teoh and Wong (2002) also find evidence of pre-IPO earnings manipulation.

lower underpricing for firms with more frequent disclosures prior to the IPO. An exception is Internet firms, where this relation is reversed. They also find that more frequent disclosure ameliorates adverse selection in the aftermarket, with lower bid-ask spreads and greater depths for firms that disclose more frequently pre-IPO. Finally, Leone et al. (2007) find that IPO underpricing decreases when issuers disclose more specific information in the "uses of proceeds" section of the prospectus. Collectively, these studies, all of which focus on IPOs in a single country, suggest that accounting disclosures influence underpricing.

Our study extends this analysis to a multi-country setting to determine if earnings quality at the country level influences underpricing costs borne by firms going public in different markets. In so doing, we add to the very limited evidence on the determinants of cross-country underpricing variation as well as contribute to another strain of the literature on the quality and value of accounting information in different countries. For example, DeFond et al. (2007) find that earnings announcements are more informative in countries with better overall earnings quality. Bhattacharya et al. (2003) study the extent to which cross-country variation in the quality of accounting information influences the cost of capital and trading volume in international equity markets. Using data from 34 countries covering 1985–1998, they measure three dimensions of earnings opacity for each country—earnings aggressiveness, loss avoidance, and earnings smoothing. They find robust evidence that an increase in overall earnings opacity leads to an increase in the required return demanded by shareholders and a decrease in trading activity. In a similar vein, Leuz et al. (2003) examine earnings management practices in 31 countries and find that firms engage in greater earnings management in countries with weaker investor protections.[4]

These studies and others suggest that the quality of earnings information available to outside investors influences information asymmetries in the market. Given the central role of asymmetric information in theories of IPO underpricing, we anticipate that differences in earnings quality across countries will influence the underpricing costs that firms going public in different countries face. Recognizing that earnings quality is a multi-faceted concept and that no single measure can capture all aspects of earnings quality, we employ a wide range of earnings quality measures drawn from the extant literature and study their association with firm-level underpricing. The measures we use are designed to capture elements of earnings quality such as earnings smoothing, loss avoidance, and earnings aggressiveness, among others. Below, we construct and use the earnings quality measures in Leuz et al. (2003) and Bhattacharya et al. (2003) to test our earnings quality hypothesis.

III. SAMPLE CONSTRUCTION AND DESCRIPTIVE STATISTICS

Sample Construction

We begin our sample construction by retrieving each IPO with a valid offer price and offer size for all countries reported in the Thomson Financial SDC Platinum New Issues database from 1998 through 2008. We exclude financial firms, rights offerings, unit offerings, closed-end funds, trusts, limited partnerships, and depository receipts. This initial extraction from SDC results in 13,390 IPOs from around the world. Next we attempt to match each IPO firm with the Datastream database to collect the secondary market prices that we need to calculate IPO initial returns. This matching process is done in two steps: (1) we use the SEDOL firm identifier included in both the SDC and Datastream databases, and (2) we hand-match any firm that does not have a SEDOL reported in both databases. We are able to match 74.4 percent of the SDC IPOs using the SEDOL and an additional 16.2 percent by hand, for a total of 12,122 IPOs matched with Datastream.

---

[4] Bushman and Piotroski (2006) construct earnings quality measures based upon accounting conservatism and link those measures to cross-country variation in legal and political institutions. See also Hung (2000) and Ball et al. (2003).

We next drop 950 IPOs that do not have a valid "first-day" secondary market closing price in Datastream. We define a valid secondary market closing price as a price observation in Datastream with positive trading volume (i.e., a transaction price). The first-day secondary market closing price is the first valid closing price that occurs within −3 to +60 days of the SDC IPO issue date.[5] From the remaining 11,172 IPOs, we drop an additional 161 deals due to incomplete or missing data for our earnings quality measures or control variables.

For 11,011 IPOs from around the world, we calculate the initial return as the first-day secondary market closing price divided by the IPO offer price, minus 1.[6] To eliminate the impact of outliers, we trim our sample by removing the top and bottom 1 percent based on initial returns.[7] Finally, we exclude countries with fewer than five IPOs during our sample period. These steps result in a final sample of 10,783 IPOs listed in 37 countries. Table 1 summarizes the sample selection process.

Although common in the IPO literature, we do not impose a minimum offer price restriction. For example, Ritter (1991) evaluates U.S. IPOs with a minimum offer price of $1 to mitigate the bid-ask bounce effect. However, imposing this filter would not only greatly reduce the number of IPOs in many countries, but it would also eliminate some countries entirely. Applying a $1 minimum offer price (converting local currency to U.S. dollars based on the exchange rate as of the IPO date) eliminates over one-third of the sample events. Thus, the main analysis presented here imposes no minimum offer price, but we do verify that our results are unaffected by the inclusion of IPOs with very low offer prices.

TABLE 1

Sample Construction

IPO Sample Selection Procedure	Number of IPOs Dropped	Remaining IPOs
(1) Pull IPOs for all countries from SDC New Issues database.		13,390
(2) Drop IPOs that cannot be matched with the Datastream database.	1,268	12,122
(3) Drop IPOs that do not have a valid first-day secondary market closing price.	950	11,172
(4) Drop IPOs from countries not covered by LLSV antidirector rights index (except for IPOs from China).	140	11,032
(5) Drop IPOs from countries for which earnings quality measures cannot be calculated for the five years prior to the IPO.	21	11,011
(6) Drop IPOs with an initial return in the 1st or 99th percentile.	223	10,788
(7) Drop IPOs from countries with less than five IPOs.	5	10,783

This table presents the sample construction process for the entire sample of 10,783 IPOs listed in 37 countries.

[5] In 39 cases, representing less than 0.4 percent of our sample, there is an apparent error in the SDC IPO date because the first closing price of the firm found in Datastream occurs one, two, or three days prior to the SDC IPO date. For France, Greece, and Taiwan we use the tenth valid price, rather than the first, because in these countries secondary prices are initially constrained by daily volatility limits for a few days following the IPO.

[6] When calculating initial returns, closing and offer prices are in the IPO firm's local currency.

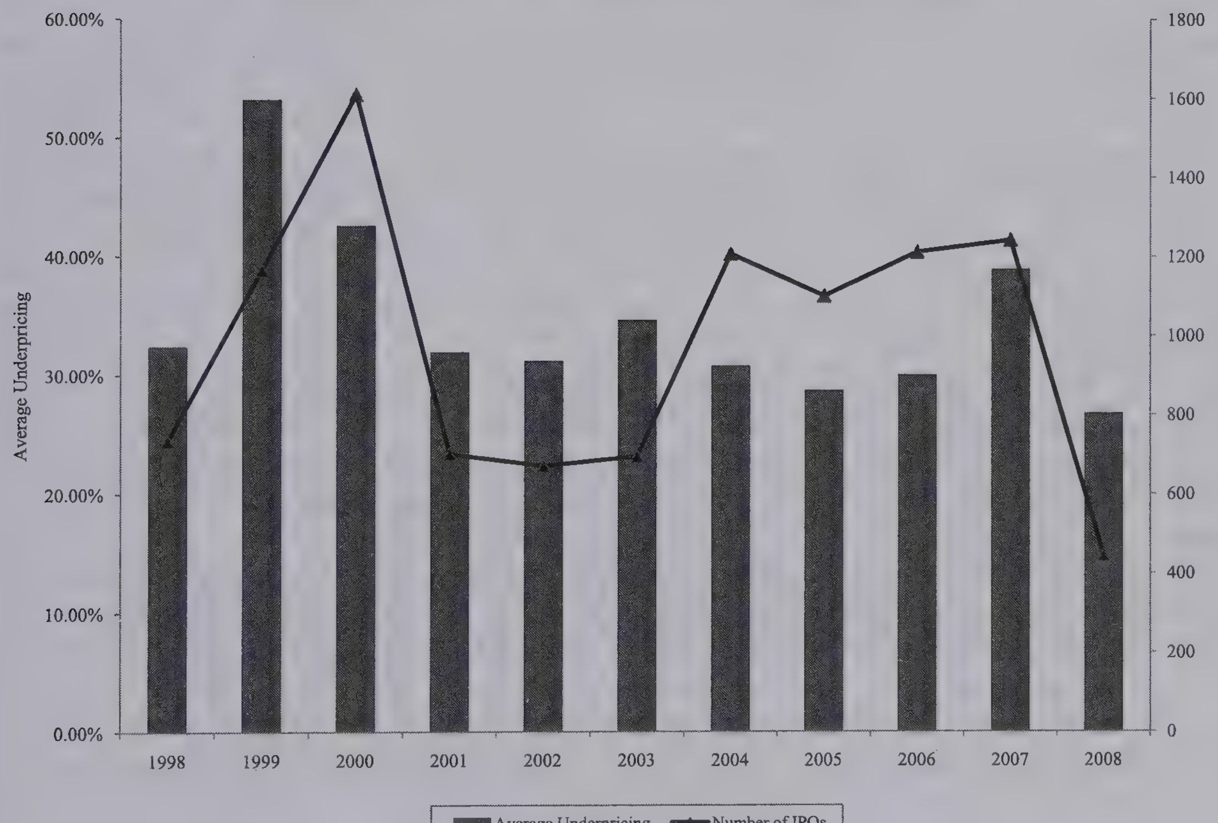
[7] Initial returns at the 1st and 99th percentiles cutoffs are −36 and 388 percent, respectively.

Descriptive Statistics

Figure 1 displays the average IPO underpricing and number of IPOs by year. The figure clearly illustrates the slowdown in equity offerings following the decline in equity markets in 2000. IPO volume falls more than 56 percent from 2000 to 2001, then remains relatively flat for 2002 and 2003. The number of new issues rebounds sharply to 1,206 deals in 2004 and remains above 1,100 issues per year through 2007. The worldwide financial crisis of 2008 had a significant negative impact on IPO issuance worldwide, with only 444 firms going public. Consistent with the "hot issues market" phenomenon (e.g., Ritter 1984), average underpricing is positively correlated with yearly IPO volume (correlation = 0.51).

Coffee (1999, 2002) suggests that firms list abroad to bond themselves to foreign listing standards. For example, firms listing in the U.S. subject themselves to SEC oversight, agree to meet generally accepted accounting principles (GAAP), and face the scrutiny of financial intermediaries involved in the security markets. While most of our IPOs originate and list in the same country, some companies choose to list outside their home country. We exclude depositary receipts from our sample, but retain firms that list their shares directly in a foreign market. Most of these

FIGURE 1
Average Underpricing and Number of IPOs by Year



Bar heights show the average underpricing by year. Line points show the number of IPOs by year.

firms list in the U.S. or the U.K. Because listing abroad can bond management to the listing country's standards, the country where the firm lists is the relevant location for this study.[8]

Table 2 shows the IPO volume, aggregate gross IPO proceeds, average underpricing, and the average value of each earnings quality variable for each country in our sample. Not surprisingly, the U.S. has the most IPOs in the sample, followed closely by Japan, the U.K., and Australia. The aggregate gross proceeds for the entire sample is about $1.12 trillion, of which the U.S. represents about 29 percent. Table 2 confirms that IPO firms are underpriced, on average, in every country, and that average underpricing varies widely across countries. First-day returns range from 120.7 percent in China to less than 2 percent in Argentina.

For each country and each year, we calculate several earnings quality measures, described in more detail in the next section. One such measure is a country's aggregate earnings management (aggregate EM) score, which is constructed such that a higher score implies greater earnings management, and therefore lower quality earnings. The average aggregate EM scores indicate that the countries with the least (most) earnings management are Australia, South Africa, and the U.S. (China, Indonesia, and Taiwan). An alternative earnings quality measure is earnings opacity. The average earnings opacity scores indicate that the countries with the most transparent (opaque) earnings are Norway, the U.S., and Canada (China, India, and Hong Kong).[9] At the bottom of Table 2 we report the simple correlations between the average IPO underpricing and earnings quality measures. The positive correlations are consistent with our earnings quality hypothesis, which predicts greater underpricing in countries with lower quality earnings.

Figure 2 offers preliminary evidence that earnings quality influences underpricing. Figure 2 groups countries into earnings quality quartiles based on their average yearly aggregate earnings management score for the sample period. Countries with the highest aggregate EM are in the lowest earnings quality quartile and those with the lowest aggregate EM are in the highest earnings quality quartile. Figure 2 reports the average underpricing and average aggregate EM across IPOs within an earnings quality quartile. Average underpricing is almost 77 percent higher (or over 21 percentage points higher) in the lowest earnings quality quartile compared to the highest quartile. Across earnings quality quartiles, the simple correlation between average IPO underpricing and average aggregate EM is 76 percent. Of course, many differences in the IPO process and in the types of firms going public exist across countries, so in the next section we test for a link between earnings quality and IPO underpricing while holding constant deal-level and other country-level characteristics.

Table 3 provides descriptive statistics for the variables used in our regression analysis, starting with various earnings quality measures drawn from Leuz et al. (2003) and Bhattacharya et al. (2003). For these measures, we provide up-to-date, yearly calculations (described below) for each of our sample countries and scale these measures such that a higher value indicates lower earnings quality.

The unit of observation in our regressions is the IPO firm, and the dependent variable is the first-day return. The sample mean initial return equals 36.5 percent, which is roughly double the 18.7 percent mean reported by Loughran and Ritter (2004) for 6,391 IPOs issued in the U.S. between 1980 and 2003. The cross-sectional standard deviation of initial returns is 61.2 percent, indicating large variations in IPO returns. Unquestionably, some of this variation can be explained by deal-specific and market-specific factors not related to earnings quality, so we use the remaining variables listed in Table 3 to control for these effects.

---

8 Unreported robustness tests confirm that our results are not sensitive to the exclusion of firms that choose to list in a foreign country.

9 To provide a check of our earnings quality variables against those published in other studies, we compared our calculations to those published in Bhattacharya et al. (2003, Table 1). For comparable variables, the correlation between our measures and theirs (which were taken from a slightly different time period) is 0.87.

TABLE 2

Country-Level Descriptive Statistics

Country	n	Aggregate Gross Proceeds (U.S.$ MM)	Average Underpricing	Average EM1	Average EM2	Average EM3	Average EM4	Average Aggregate EM	Average Earnings Aggressiveness	Average Earnings Opacity
Argentina	5	247.73	1.43%	−0.73	0.78	0.48	0.89	14.60	−0.06	3.88
Australia	1,009	26,659.45	23.06%	−0.74	0.81	0.47	0.62	7.46	−0.04	4.10
Austria	31	7,278.92	6.10%	−0.32	0.94	0.69	0.75	31.44	−0.06	4.73
Belgium	61	14,515.28	10.19%	−0.37	0.93	0.66	0.78	29.75	−0.07	4.58
Brazil	46	16,982.60	6.17%	−0.57	0.90	0.51	0.67	14.17	−0.02	5.99
Canada	421	13,138.84	48.61%	−0.60	0.88	0.51	0.66	14.13	−0.05	3.86
China	686	58,730.82	120.72%	−0.37	0.96	0.70	0.97	35.87	−0.01	8.03
Denmark	27	2,341.67	7.09%	−0.40	0.89	0.57	0.76	24.12	−0.04	4.65
Finland	43	6,063.64	26.14%	−0.47	0.91	0.46	0.94	24.01	−0.04	5.57
France	495	71,490.92	15.84%	−0.39	0.97	0.58	0.76	28.55	−0.04	5.54
Germany	366	59,422.92	36.63%	−0.43	0.95	0.64	0.79	27.68	−0.06	4.97
Greece	109	5,272.99	58.33%	−0.29	0.90	0.63	0.93	33.17	0.01	6.00
Hong Kong	583	79,494.15	16.69%	−0.36	0.96	0.75	0.74	33.08	−0.02	6.60
India	169	16,974.68	40.21%	−0.47	0.98	0.45	0.83	26.71	−0.02	7.58
Indonesia	60	5,260.06	39.32%	−0.31	0.95	0.83	0.75	35.26	−0.04	5.64
Ireland	7	2,735.77	3.85%	−0.63	0.95	0.42	0.84	19.75	−0.03	6.03
Israel	9	251.67	24.28%	−0.40	0.94	0.70	0.78	29.94	−0.03	5.58
Italy	156	50,044.99	10.46%	−0.28	0.93	0.67	0.77	32.13	−0.04	5.49
Japan	1,259	85,327.98	55.85%	−0.37	0.94	0.63	0.79	30.37	−0.03	6.40
Malaysia	309	5,944.56	34.32%	−0.44	0.91	0.78	0.74	28.51	−0.02	6.27
Mexico	11	3,113.55	7.69%	−0.46	0.84	0.43	0.71	15.61	−0.03	5.30
The Netherlands	46	14,118.38	28.48%	−0.42	0.92	0.51	0.74	22.67	−0.05	4.60
New Zealand	36	2,191.18	10.45%	−0.57	0.60	0.39	0.69	9.53	−0.04	4.71
Norway	96	14,112.62	3.74%	−0.59	0.32	0.66	0.60	13.50	−0.06	3.30
Philippines	22	1,831.36	12.32%	−0.43	0.84	0.76	0.69	27.59	−0.04	5.19
Portugal	10	6,242.68	15.95%	−0.28	0.94	0.66	0.80	33.88	−0.05	5.30
Singapore	462	14,389.05	27.08%	−0.39	0.96	0.80	0.71	32.22	−0.03	6.18

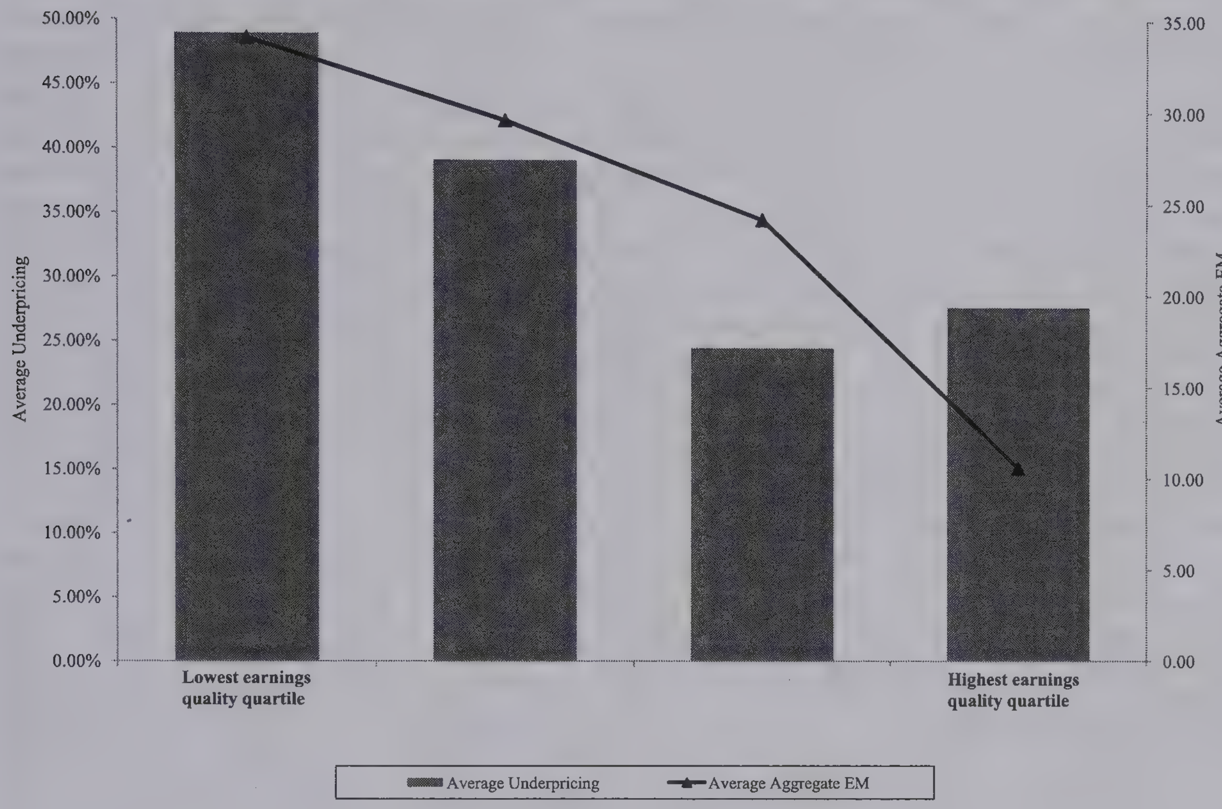
(continued on next page)

TABLE 2 (continued)

Country	n	Aggregate Gross Proceeds (U.S.$ MM)	Average Underpricing	Average EM1	Average EM2	Average EM3	Average EM4	Average Aggregate EM	Average Earnings Aggressiveness	Average Earnings Opacity
South Africa	8	1,082.68	8.53%	−0.59	0.79	0.42	0.53	7.94	−0.04	4.23
South Korea	517	21,252.00	46.63%	−0.37	0.92	0.64	0.75	28.04	−0.04	5.70
Spain	38	27,860.74	14.76%	−0.36	0.97	0.54	0.86	30.64	−0.04	6.15
Sweden	76	17,718.82	9.21%	−0.60	0.90	0.52	0.65	14.33	−0.04	4.91
Switzerland	69	24,351.77	11.49%	−0.41	0.93	0.52	0.83	25.81	−0.05	5.46
Taiwan	547	10,382.86	22.96%	−0.36	0.97	0.79	0.75	34.01	−0.03	6.49
Thailand	142	11,492.59	17.31%	−0.41	1.00	0.65	0.74	31.96	−0.05	5.95
Turkey	16	3,870.52	7.22%	−0.69	0.84	0.67	0.80	20.83	0.03	5.64
U.K.	1,044	91,006.69	17.69%	−0.59	0.88	0.49	0.69	14.92	−0.05	4.52
U.S.	1,792	325,125.95	33.90%	−0.57	0.68	0.46	0.63	8.67	−0.04	3.63
Correlation with average underpricing				0.30	0.25	0.23	0.40	0.36	0.31	0.49
p-value				(0.075)	(0.139)	(0.177)	(0.014)	(0.030)	(0.060)	(0.002)

This table presents country-level descriptive statistics for the entire sample of 10,783 IPOs. n is the number of sample IPOs listing in the country over the 1998–2008 sample period. Aggregate gross proceeds is the sum of CPI-adjusted offer value in 2008 U.S. dollars for the country of listing. Underpricing is the first-day secondary market closing price divided by the final offer price, minus 1. EM1 is the median ratio in country i of the firm-level standard deviations of operating earnings over the cash flow from operations (both scaled by lagged total assets), multiplied by −1. EM2 is the cross-sectional correlation in country i between the change in accruals and change in cash flows from operations (both scaled by lagged total assets), multiplied by −1. EM3 is the median ratio in country i of the absolute value of accruals over the absolute value of cash flow from operations. EM4 is the ratio in country i of the number of firms reporting small profits over the sum of the number of firms reporting small losses and small profits. A small profit (loss) is defined as a value of net earnings scaled by lagged total assets in the range [0, 0.01] ([−0.01, 0)). Aggregate EM is the average country i ranking across the following four earnings management measures: EM1, EM2, EM3, and EM4. Earnings aggressiveness is the median ratio in country i of total accruals over the lagged total assets. Earnings opacity is the average country i decile ranking across the following three earnings management measures: EM2, EM4, and earnings aggressiveness. The final two rows report the correlation between the average earnings quality measure and average underpricing and the p-value from a test of the null hypothesis that the correlation is equal to 0.

FIGURE 2
Average Underpricing and Average Aggregate EM by Earnings Quality Quartiles



Sample countries are sorted into quartiles based on their average aggregate EM score over the 1998–2008 sample period. Countries with the highest aggregate earnings management are in the lowest earnings quality quartile and those with the lowest aggregate earnings management are in the highest earnings quality quartile. Bar heights reflect the average underpricing across all IPOs within a quartile. Line points reflect the average aggregate EM score across all IPOs within a quartile.

Because the extent to which an IPO is underpriced may be influenced by the quality of the underwriter, we create an indicator variable to identify top-tier underwriters. Underwriters listed in the top 25 of SDC's global league tables for the issue year are coded as top-tier underwriters. Table 3 indicates that 22.5 percent of our IPOs employ a top-tier underwriter.[10]

In some countries, underwriters can engage in price support once trading begins. Underwriters may have incentives to engage in price stabilization, particularly when an IPO's market price begins to fall below the IPO offer price. If underwriters engage in price support to limit the occurrence of negative initial returns, then the first-day returns distribution will have a higher mean than it would have in the absence of price support. We do not have detailed data on the regulations and practices with respect to price stabilization in all countries, so we attempt to control for price-stabilization activity in two ways. First, we construct a country-level proxy for

[10] In unreported tests, we controlled for underwriter reputation using country-level or global market share measures. These measures are consistent with Megginson and Weiss (1991), who use market share to proxy for underwriter reputation. Our results are broadly similar when these measures are used to proxy for underwriter reputation.

TABLE 3

IPO Sample Descriptive Statistics

Variable	n	Mean	Std. Dev.	Minimum	Maximum
EM1	10,783	−0.480	0.139	−0.844	−0.173
EM2	10,783	0.876	0.150	−0.402	1.000
EM3	10,783	0.592	0.131	0.275	0.996
EM4	10,783	0.728	0.103	0.367	1.000
Aggregate EM	10,783	22.388	11.145	4.750	45.000
Earnings aggressiveness	10,783	−0.036	0.015	−0.100	0.206
Earnings opacity	10,783	5.326	1.392	2.867	8.667
Initial return	10,783	0.365	0.612	−0.360	3.862
Top-tier underwriter	10,783	0.225	0.418	0.000	1.000
Price stabilization	10,783	0.007	0.019	−0.074	0.125
IPO activity	10,783	0.045	0.029	0.001	0.146
Market return	10,783	0.030	0.112	−0.488	1.132
Stock market turnover ratio	10,783	1.109	0.638	0.074	2.946
Index of economic freedom	10,783	72.568	9.738	44.858	91.367
Antidirector rights index	10,783	3.834	1.199	0.000	5.000
Offer size ratio	10,783	1.000	4.861	0.000	351.819
Stock return volatility	10,708	0.048	0.040	0.000	1.571
Integer offer price	10,783	0.500	0.500	0.000	1.000
Bookbuilt	10,198	0.612	0.487	0.000	1.000
Firm commitment	10,707	0.560	0.496	0.000	1.000
Equity carve-out	10,699	0.034	0.181	0.000	1.000

This table presents descriptive statistics for the entire sample of 10,783 IPOs. EM1 is the median ratio in country *i* of the firm-level standard deviations of operating earnings over the cash flow from operations (both scaled by lagged total assets) multiplied by −1. EM2 is the cross-sectional correlation in country *i* between the change in accruals and change in cash flows from operations (both scaled by lagged total assets) multiplied by −1. EM3 is the median ratio in country *i* of the absolute value of accruals over the absolute value of cash flow from operations. EM4 is the ratio in country *i* of the number of firms reporting small profits over the sum of the number of firms reporting small losses and small profits. A small profit (loss) is defined as a value of net earnings scaled by lagged total assets in the range [0, 0.01] ([−0.01, 0)). Aggregate EM is the average country *i* ranking across the following four earnings management measures: EM1, EM2, EM3, and EM4. Earnings aggressiveness is the median ratio in country *i* of total accruals over the lagged total assets. Earnings opacity is the average country *i* decile ranking across the following three earnings management measures: EM2, EM4, and earnings aggressiveness. Initial return is the secondary market closing price divided by the final offer price minus 1. Top-tier is an indicator variable set to 1 for IPOs underwritten by an investment bank appearing in the top 25 of SDC's league tables in the issue year, and 0 otherwise. Price stabilization is the difference in the number of IPOs with small positive first-day returns (greater than 0 and less than or equal to 1 percent) and the number of IPOs with small negative first-day returns (less than 0 and greater than or equal to −1 percent) divided by the total number of IPOs issued in the country of listing. IPO activity is the ratio of the total number of IPOs in the issue year divided by the number of Datastream-listed equities for the country of listing as of 2008. Market return is the return on the Datastream index for the country of listing over the three months preceding the offering. Stock market turnover ratio equals the ratio of the total value of shares traded to aggregate market capitalization and reported in Beck et al. (2000). Index of economic freedom is a product of The Heritage Foundation and the *Wall Street Journal* and is an aggregate measure covering the following ten freedoms: business, trade, monetary, freedom from government, fiscal, property rights, investment, financial, freedom from corruption, and labor. Antidirector rights index measures shareholder rights by considering the following issues: (1) vote by mail, (2) shares not blocked or deposited prior to shareholder meetings, (3) cumulative voting in director elections, (4) oppressed minority mechanisms, (5) pre-emptive rights to new issues, and (6) minimum capital requirements and is reported in La Porta et al. (1998) and Allen et al. (2005). Offer size ratio is the CPI-adjusted offer value in millions of U.S. dollars divided by the average of this value across all IPOs in the country of listing. Stock return volatility is the standard deviation of returns over the first 30 calendar days following the offering. Indicator variables are set equal to 1 for integer offer price, bookbuilt, firm commitment, and equity carve-out deals.

price-stabilization activity. If price stabilization is widespread, then we expect to see an unusually large probability mass in the distribution of first-day returns at and just to the right of zero and an unusually small probability mass immediately to the left of zero. Therefore, for each country, we calculate the difference in the number of IPOs with initial returns between 0 and 1 percent and the number of IPOs with initial returns between 0 and −1 percent, and then divide the difference by the total number of IPOs. The more prevalent is price stabilization in a given market the higher should be this ratio, and the higher should be average underpricing. Table 3 reports a mean price stabilization ratio of 0.007, indicating a slightly greater incidence of small positive initial returns than small negative returns. A second approach to control for price stabilization exploits the fact that underwriters typically provide price support for a very limited time. Because stabilization activities are short-lived, the effects of stabilization activities on the IPO returns distribution should diminish over time. In unreported tests we calculate the IPO initial return based on the market price roughly one month (22 trading days) after the first trading day, and test to see if our regression results are robust to this change.

As suggested in numerous studies, including Ritter (1984), underpricing tends to be higher when IPO volume is high and when overall stock market returns are high. We include two variables to control for these "hot market" effects. First, our IPO activity measure equals the number of IPOs in a given country in each year divided by the total number of listed equities in Datastream for that country in 2008. Therefore, this measure takes the same value for all IPOs from a single country in a particular year, but within a country it varies across time, and within a single year it varies across countries. Second, for each IPO, we calculate the return on the Datastream market index in the three months leading up to the offer. Consequently, two firms will share the same market return value only if they go public at the same time in the same country.

Ellul and Pagano (2006) suggest that IPOs in less liquid markets will exhibit larger initial returns. Higher underpricing compensates IPO investors for the illiquidity risk that they bear. To control for differences in liquidity across national markets, we include a country-level stock market turnover ratio, which is defined by Beck et al. (2000) as the ratio of total value of shares traded to market capitalization.[11] Hence, this measure captures a stock market's liquidity relative to its size, and we expect lower initial returns in more liquid markets.

The climate for entrepreneurship varies from one country to another due to a variety of factors such as tax rates, government regulations, corruption, and many others. To control for some of these differences, we include the Index of Economic Freedom, which is produced by the Heritage Foundation and the *Wall Street Journal*. The index is an aggregate of several different measures of economic freedom, including: (1) the ease of starting a new business, (2) the level of trade barriers, (3) the top marginal income tax rate, (4) government spending as a percentage of GDP, (5) the inflation rate, (6) the ease with which foreigners can invest in a given country, (7) the extent of government control of the banking system, (8) legal protection of property rights, (9) corruption, and (10) the ability of a business to hire and fire workers without restrictions.[12] A higher index value indicates greater economic freedom.

Because of the literature spawned by the work of La Porta et al. (2000) illustrating the importance of various investor protections on a range of capital market outcomes, we also attempt to control for variation in governance across our sample countries. The index of economic free-

---

[11] Beck et al. (2000) report the ratio of total value of shares traded to market capitalization for the period 1999–2004. Since our sample period is 1998–2008, we take the average of all years reported for a country by Beck et al. (2000) as our measure of country-level stock market turnover.

[12] Moeller and Schlingemann (2005) use this index to proxy for the restrictiveness of the institutional environment related to corporate acquisitions, and they find that in cross-border acquisitions, bidder returns are higher when targets come from countries with greater freedom.

dom, given its constituent parts, should provide one measure of control for differences in country-level governance. A second control variable, aimed at holding constant the effects of governance, is the antidirector rights index from La Porta et al. (1998). The antidirector rights index is based on a collection of legal and regulatory variables related to a nation's stance regarding the protection of shareholders. Boulton et al. (2010) report a positive relation between the antidirector rights index and IPO underpricing, which they suggest is a cost incurred by insiders to maintain control in countries with legal systems that offer stronger protections to investors.

Most IPO underpricing studies include measures designed to capture information asymmetries, including the natural logarithm of the offer size. Because the amount of money that IPO firms raise varies widely across countries, we also construct an offer size measure that captures the size of an IPO relative to other deals in the same country, which we report in Table 3. This offer size ratio equals the offer size for a particular IPO divided by the mean offer size for IPOs in a given country.[13] We obtain broadly similar results with both absolute and relative deal-size measures, so in Table 3 and in subsequent regressions we simply include the offer size ratio for each IPO. In addition to offer size, we also include a measure of stock return volatility to control for information asymmetry. Our stock return volatility measure is the standard deviation of returns over the first 30 (calendar) days following the IPO.

Bradley et al. (2004) report that IPOs with integer prices experience higher underpricing. To explain this pattern, they propose a "negotiation hypothesis" in which underwriters and issuers bargain over finer offer price increments as the uncertainty surrounding firm value declines. Thus, for IPOs that are particularly difficult to value and, hence, should have higher underpricing, offer prices tend to fall on integers. Fifty percent of our IPOs have an integer price, which compares to 76 percent of U.S. IPOs as reported in Bradley et al. (2004). We include in our regression an indicator variable equal to 1 when the offer price is an integer.[14]

Sherman (2005) notes that methods for taking firms public worldwide are converging toward the traditional U.S. bookbuilding approach, and she argues that bookbuilt offers are associated with lower underpricing than offers conducted via auctions. Sherman (2005) argues that bookbuilding reduces the risk faced by IPO issuers and investors, and therefore should lead to less underpricing. However, she also notes that bookbuilding affords issuers discretion over information expenditures, which she suggests can lead to either more or less underpricing depending on the issuer's preferences. Ritter (1987) provides evidence that firm commitment IPOs are underpriced less than best efforts IPOs. Sixty-one percent of the firms in our sample are taken public through a bookbuilt offering, while 56 percent of the deals are firm commitment offerings. Deals that are both firm commitment and bookbuilt represent 31.6 percent of the sample. We include indicator variables for bookbuilt deals and for firm commitment offerings. Schipper and Smith (1986) and Prezas et al. (2000) provide evidence that equity carve-outs are underpriced less than original IPOs. To capture this effect, we include an indicator variable for carve-outs.

IV. EARNINGS QUALITY

Country-Level Earnings Quality and IPO Underpricing

In this section, we ask whether earnings quality affects underpricing. We calculate two different sets of well-established earnings management measures that are intended to capture the variation in the quality of earnings information across our sample countries. Our premise is that information asymmetry and earnings quality are inversely related. The basic tenet of our earnings

---

[13] All offer sizes are measured in millions of CPI-adjusted 2008 U.S. dollars. CPI data are from the U.S. Bureau of Labor Statistics website. SDC reports offer size in U.S. dollars.

[14] In Japan all IPOs are priced on an integer. For Japanese IPOs, our integer indicator equals 1 if the offer price (in yen) is perfectly divisible by 100.

quality hypothesis is that managers can take various actions to obscure the true earnings distribution, and that these actions increase information asymmetries and underpricing. To construct our earnings quality measures, we use financial statement data on firms in each country that are already publicly traded. In this way, we attempt to characterize the earnings management practices that are common in each country rather than capturing earnings management behavior for each IPO firm. We use firm-level data from Compustat Global over the years 1993 to 2007 for all nonfinancial firms with a minimum of three consecutive years of accounting data. Each country included in the sample has a minimum of 200 firm-year observations from Compustat Global. Each earnings management measure is constructed annually over the 1998–2008 sample period using accounting data from the prior five years. Thus, our earnings quality variables vary over time and across countries.

Our first set of earnings quality measures come from Leuz et al. (2003). They construct four country-level earnings management measures. The first of these, denoted EM1, is an earnings smoothing measure equal to the median ratio in country i of the firm-level standard deviations of operating earnings over the cash flow from operations (both scaled by lagged total assets). Their second measure, EM2, also attempts to capture earnings smoothing behavior. EM2 equals the cross-sectional correlation in country i in year t between the change in accruals and the change in cash flows (both scaled by lagged total assets). We transform our EM1 and EM2 measures by multiplying their values by -1, so that for EM1, EM2, and all other earnings quality measures, higher values correspond to more aggressive earnings management. Thus, our earnings quality hypothesis predicts a positive coefficient on the earnings quality measures in underpricing regressions, meaning that more earnings management leads to greater information asymmetry and higher underpricing.

The third measure, EM3, is the median value in each country of the absolute value of firms' accruals scaled by the absolute value of cash flow from operations. A higher value of EM3 indicates greater earnings management. The final metric, EM4, measures loss avoidance behavior. EM4 equals the ratio of the number of firms reporting small profits over the sum of the number of firms reporting small losses and small profits. In this context, "small" means a ratio of net income to assets of plus or minus 1 percent. The intuition for this measure is that if managers manipulate earnings to avoid showing losses, then there will be a missing probability mass in the earnings distribution just to the left of breakeven, and a higher-than-expected frequency of firms reporting earnings just above zero (see Degeorge et al. 1999). Therefore, the higher is the loss avoidance ratio, the greater is the incidence of loss avoidance behavior in a given country and the more opaque are the country's earnings figures. Finally, Leuz et al. (2003) calculate an overall earnings management figure, aggregate EM, for each country by ranking each country on each of the four earnings management measures and then taking the average ranking, where higher rankings signify more earnings management.

In addition to calculating two earnings management measures that replicate measures from Leuz et al. (2003), Bhattacharya et al. (2003) construct two additional earnings quality measures that focus on aspects of earnings opacity. The first unique measure is their earnings aggressiveness measure, defined as the tendency to accelerate the recognition of gains and delay the recognition of losses. Earnings aggressiveness for country i in year t is equal to the median value of the ratio of total accruals divided by lagged assets. A higher value of this ratio implies more aggressive (and more opaque) earnings. The second unique measure is the Bhattacharya et al. (2003) earnings opacity measure. They construct a single aggregate earnings opacity measure for each country based on a ranking methodology. Specifically, deciles are constructed for each of three earnings quality measures and then the decile ranks are averaged across the three measures to arrive at an overall earnings opacity ranking for each country. Countries that earn a higher average ranking

have more opaque earnings. In the regression analyses that follow, we estimate the relation between each earnings quality measure and IPO underpricing.

Table 4 reports regressions that examine the relation between earnings quality and underpricing on a country-year basis. The dependent variable in the Table 4 regressions is the average IPO initial return in a given country and year. The control variables are country-year averages across all IPOs. There are 320 unique country-year combinations in our sample. The primary variables of interest in Table 4 are our aggregate EM measure in Model 1 and earnings opacity score in Model 2, where higher values are indicative of more earnings management. Our earnings quality hypothesis predicts a positive relation between our earnings management measures and IPO initial returns.

The Table 4 results support the prediction of our earnings quality hypothesis. Average underpricing is positively related to both aggregate EM and earnings opacity. The aggregate EM coefficient in Model 1 (0.004) implies that a one standard deviation improvement in a country's average aggregate EM measure is related to a reduction in average underpricing of 4.45 percentage points. The control variables in Table 4 are broadly consistent with expectations. The positive coefficients for IPO activity and market return are consistent with the "hot markets" effect. We also find that underpricing is positively correlated with the level of protection afforded to minority shareholders (as captured by the antidirector rights index) and stock return volatility. Underpricing is negatively correlated with the index of economic freedom.

TABLE 4

Country-Year Underpricing Regressions on Earnings Quality

	Model 1	Model 2
Intercept	0.392	0.105
Aggregate EM	0.004**	
Earnings opacity		0.063 ***
Top-tier underwriter	0.060	0.064
Price stabilization	−0.219	−0.396
IPO activity	1.839***	1.675 ***
Market return	0.740***	0.757 ***
Stock market turnover ratio	0.026	0.022
Index of economic freedom	−0.010***	−0.008 ***
Antidirector rights index	0.041***	0.026 *
Offer size ratio	0.000	−0.001
Stock return volatility	3.558***	3.777 ***
Integer offer price	−0.013	0.012
Bookbuilt	−0.093*	−0.077
Firm commitment	0.044	0.005
Equity carve-out	−0.016	−0.035
Adjusted R^2	27.64%	29.90%
Number of observations	320	320

*, **, *** Denote significance of the coefficient at the 10, 5, and 1 percent levels, respectively.

This table presents OLS regressions of IPO underpricing on country-level earnings quality measures. The dependent variable is the average IPO initial return measured at the country-year level. Initial return is calculated by dividing the secondary market closing price by the final offer price minus 1. The dependent variables are averaged at the country-year level and are defined in the notes to Table 3. Regressions control for industry composition. Industry classifications reflect those reported by Dyck and Zingales (2004).

To further examine the association between country-level earnings quality and IPO underpricing, in Table 5 we run pooled cross-sectional regressions of firm-level IPO underpricing on country-level measures of earnings quality and a variety of country- and firm-level controls.[15] In Models 1–4, we estimate regressions with each Leuz et al. (2003) earnings measure separately. In Model 5, we consider the aggregate earnings management measure. To the extent that earnings quality is related to information asymmetry, we expect poorer earnings quality to be associated with higher underpricing, on average.

Recall that for EM1, EM2, EM3, EM4, and the overall aggregate EM variable, a higher value means more earnings management. Therefore, our earnings quality hypothesis predicts positive coefficients for all these measures of earnings quality. The results from Models 1–4 provide support for the earnings quality hypothesis. Each of the earnings management measures enters the regression with a positive coefficient, as expected, although only two of the four coefficients are significant. When it is common practice in a country for managers to engage in activities to manipulate earnings, IPO underpricing is higher. Given the tendency of the individual earnings management variables to link to underpricing as expected, it is not surprising that, in Model 5, the aggregate EM score also indicates that poor earnings quality is associated with higher IPO underpricing.

In Table 5, Models 6 and 7, we examine the influence of the Bhattacharya et al. (2003) earnings aggressiveness and earnings opacity measures on IPO underpricing. These measures are constructed such that a higher value implies more opaque earnings. Therefore, we expect positive signs in the regressions. Indeed, both coefficients are positive and significant, consistent with the hypothesis that underpricing is higher in countries where investors receive lower quality accounting information.[16]

Although many of the control variables in Table 5 do not show a statistically significant link to underpricing, the variables that are significant generally have the expected sign. For example, IPOs taken public after a period of high market returns or those that display greater price volatility after trading begins have higher underpricing. Consistent with Boulton et al. (2010), initial returns tend to be greater for IPOs issued in countries offering stronger protections for minority shareholders, as indicated by the positive coefficient on the antidirector rights index. As expected, equity carve-outs are underpriced less than other IPOs, on average, as are firms going public in countries with greater economic freedoms. The adjusted R^2 values indicate that our models explain between 18 and 23 percent of the variation in the international underpricing cross-section.

Economically, the earnings quality effects presented in Table 5 are quite dramatic. The results in Model 5 suggest that, for a one standard deviation increase in a firm's country-level aggregate EM measure, underpricing increases by roughly 7.8 percentage points. This increase is more than one-fifth of the sample mean initial return of 36.5 percent. As a point of reference, consider that the aggregate EM measure for India is slightly more than one standard deviation above that of the United Kingdom. Clearly, country-level earnings quality has a considerable influence on underpricing around the world.

Robustness of Country-Level Earnings Quality and IPO Underpricing

In unreported regressions, we confirm that the results presented in Tables 4 and 5 are robust to a variety of alternative specifications. In particular, we attempt to eliminate concerns regarding

---

[15] Because underpricing is likely to be correlated across IPOs within a country and across time, we construct two-way cluster-robust standard errors as outlined in Gow et al. (2010) for all firm-level IPO underpricing regressions.

[16] If we estimate the relation between underpricing and earnings quality in the presence of country-level fixed effects, the coefficients on the earnings quality measures are positive and significant about half the time.

TABLE 5

Underpricing Regressions on Earnings Quality Measures

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Intercept	1.606**	1.293*	1.282*	−0.684	1.171*	1.526**	0.699
EM1	0.455***						
EM2		0.183					
EM3			0.304				
EM4				1.796***			
Aggregate EM					0.007**		
Earnings aggressiveness						5.002***	
Earnings opacity							0.080**
Top-tier underwriter	0.044	0.050	0.047	0.076	0.054	0.052	0.066
Price stabilization	2.237	1.862	2.019	1.928	2.467	1.484	2.573
IPO activity	0.321	−0.019	0.124	0.432	0.183	0.057	0.139
Market return	0.892***	0.881***	0.875***	0.866***	0.876***	0.906***	0.868***
Stock market turnover ratio	−0.041	−0.043	−0.037	−0.004	−0.028	−0.036	−0.014
Index of economic freedom	−0.020**	−0.020**	−0.021**	−0.011**	−0.020**	−0.018**	−0.017**
Antidirector rights index	0.079**	0.062*	0.071*	0.086**	0.085*	0.040	0.069*
Offer size ratio	−0.001	−0.001	−0.001	−0.002	−0.001	−0.001	−0.001
Stock return volatility	3.949***	3.967***	4.048***	3.949***	4.027***	3.870**	4.002***
Integer offer price	−0.008	−0.004	−0.002	0.035	0.004	−0.011	0.019
Bookbuilt	−0.149	−0.136	−0.121	−0.126*	−0.135	−0.093	−0.109
Firm commitment	0.001	0.016	−0.004	0.011	−0.015	0.000	−0.049
Equity carve-out	−0.045	−0.059**	−0.052*	−0.056**	−0.051*	−0.054**	−0.058**
Adjusted R^2	18.43%	17.95%	18.05%	22.54%	18.60%	19.06%	19.57%
Number of observations	10,045	10,045	10,045	10,045	10,045	10,045	10,045

*, **, *** Denote significance of the coefficient at the 10, 5, and 1 percent levels, respectively, measured using two-way cluster-robust standard errors (country and issue year) as outlined in Gow et al. (2010).

This table presents OLS regressions of IPO underpricing on country-level earnings quality measures. The dependent variable is the IPO initial return, which is the secondary market closing price divided by the final offer price minus 1. All other variables are defined in the notes to Table 3. Regressions include industry indicators. Industry classifications reflect those reported by Dyck and Zingales (2004).

the price-stabilization activity of underwriters, which can vary across countries, minimum offer price restrictions, which screen out the smallest IPOs, and the effects of large countries, which tend to be either very transparent or very opaque.

To confirm that our results are not sensitive to variations in price-stabilization activity across countries, we replace the one-day initial return measure used in Table 5 with the IPO initial return measured as the percentage difference between the offer price and the secondary market closing price 22 trading days after the first trading date. The intuition for this alternative measure is that the effects of price stabilization dissipate over time, generally within a month of the IPO, as reported in Ruud (1993) and elsewhere. Thus, if price stabilization temporarily obscures the left tail of the IPO returns distribution, then our finding that lower quality earnings leads to higher underpricing might be the result of greater price support in countries with lower earnings quality. We find that measuring the initial return over a longer horizon does not fundamentally change our results.

In other tests, we impose increasingly stringent restrictions on our country-level sample by imposing a minimum offer price cutoff. In particular, we exclude the bottom 2, 5, 10, and 20 percent of offer prices from respective country-level samples. In terms of the U.S. distribution of offer prices, these cutoffs correspond to minimums of \$5, \$6, \$7, and \$9, respectively. Increasing the minimum offer price reduces the odds of finding spurious results driven by market microstructure effects. In all cases, the results indicate that an economically and statistically significant increase in underpricing occurs when earnings quality deteriorates, and the magnitude of this effect is similar to our estimates in all prior regressions. We also exclude each of the three largest IPO markets in our sample, the U.S., the U.K., and Japan. We do this to verify that the earnings quality results are not driven exclusively by these large markets. Excluding any of these three countries, lower earnings quality is still associated with higher underpricing.

Earnings Quality and the Certification Effect of Financial Intermediaries

The previous section presented evidence that lower earnings quality is associated with higher underpricing. In the larger context of the IPO literature, we suggest that investors face greater uncertainty regarding the values of IPO firms when earnings quality is low. One mechanism that could mitigate this uncertainty is the presence of a reputable intermediary that, in effect, certifies the earnings of IPO firms. Carter and Manaster (1990), Megginson and Weiss (1991), and Barry et al. (1990) all report results consistent with a certification effect (i.e., lower underpricing) when firms go public with the assistance of reputable underwriters or venture capital investors.[17] However, recent underpricing studies, such as Loughran and Ritter (2004), find a positive relation between underwriter quality and initial returns. Why IPOs underwritten by more prestigious investment banks are underpriced more remains an unanswered question, but in spite of this general effect, it may be possible for high-quality underwriters to mitigate the effects of poor country-level earnings quality by providing a certification effect at the firm level.

Table 6 includes interactions between our top-tier underwriter variable and mean-adjusted earnings quality measures. In Model 1, the top-tier underwriter indicator variable is interacted with the mean-adjusted aggregate EM measure, while in Model 2 the interaction is with our earnings opacity measure. If top-tier underwriters certify the earnings of IPO firms and thereby reduce the uncertainty faced by investors, then we expect the interaction terms to have the opposite sign of the relevant earnings quality measure. In both regression specifications in Table 6, the coefficient on the interaction between top-tier underwriter and earnings quality is significant and takes the opposite sign of the earnings quality coefficient.

---

[17] Due to missing data, reported results do not control for venture capital backing. Unreported tests indicate that all results are robust to controlling for venture capital backing where available.

TABLE 6

Underpricing Regressions on Earnings Quality with Underwriter Certification

	Model 1	Model 2
Intercept	1.133*	0.590
Aggregate EM	0.008***	
Earnings opacity		0.096***
Top-tier underwriter	0.042	0.048
Top-tier underwriter * earnings measure	−0.007**	−0.069*
Price stabilization	2.122	2.145
IPO activity	0.247	0.225
Market return	0.873***	0.870***
Stock market turnover ratio	−0.041	−0.027
Index of economic freedom	−0.020**	−0.016**
Antidirector rights index	0.080*	0.061
Offer size ratio	−0.001	−0.001
Stock return volatility	4.022***	4.023***
Integer offer price	0.006	0.022
Bookbuilt	−0.128	−0.098
Firm commitment	−0.009	−0.045
Equity carve-out	−0.047*	−0.054*
Adjusted R^2	18.86%	19.93%
Number of observations	10,045	10,045

*, **, *** Denote significance of the coefficient at the 10, 5, and 1 percent levels, respectively, measured using two-way cluster-robust standard errors (country and issue year) as outlined in Gow et al. (2010).

This table presents OLS regressions of IPO underpricing on country-level earnings quality measures and the interaction of underwriter reputation and earnings quality. The dependent variable is the IPO initial return, which is the secondary market closing price divided by the final offer price minus 1. All other variables are defined in the notes to Table 3. Regressions include industry indicators. Industry classifications reflect those reported by Dyck and Zingales (2004).

The coefficients on the interaction terms indicate that the less transparent a country's earnings, the more IPO issuers benefit from underwriter quality. To illustrate the economic significance of this result, consider an issuer in Italy, the country with the median average earnings opacity score (average earnings opacity score = 5.49). The coefficient on the interaction term in Model 2 of Table 6 is −0.069 and indicates that an Italian IPO firm using a top-tier underwriter would underprice nearly 1.13 percentage points less than a similar Italian IPO firm not using a top-tier underwriter. Considering that the average IPO listed in Italy is underpriced by 10.46 percent, this result implies that the presence of a top-tier underwriter reduces underpricing by more than 10 percent for new issues in Italy. The effect of a top-tier underwriter is even more pronounced for issuers from countries with higher average earnings opacity scores. Table 6, Model 2 indicates that a Chinese IPO firm (average earnings opacity score = 8.03) would underprice nearly 18.66 percentage points less when using a top-tier underwriter as opposed to an underwriter not identified as top-tier. This represents a 15 percent decrease based on the average underpricing of Chinese IPOs (120.72 percent). We obtain similar results when examining the interaction between the other proxies for earnings quality and top-tier underwriters. These results are consistent with the presence of a top-tier underwriter acting as certification for new issues and are particularly important in countries with lower earnings quality.

V. CONCLUSION

Our primary contribution is to study how country-level differences in earnings quality influence IPO underpricing. Using a wide range of earnings quality measures, we find higher IPO underpricing in countries with poorer quality earnings information, even after controlling for many country- and deal-specific characteristics. This evidence is consistent with asymmetric information explanations for underpricing. Just as other researchers have found that poor accounting information can lead to a higher cost of capital, our evidence suggests that the cost of going public rises when investors have greater difficulties interpreting financial information. However, financial intermediaries, such as investment banks, play a role in mitigating the effects of low earnings quality on IPO underpricing.

REFERENCES

Aharony, J., C. Lin, and M. P. Loeb. 1993. Initial public offerings, accounting choices, and earnings management. *Contemporary Accounting Research* 10 (1): 61–81.

———, C. J. Lee, and T. J. Wong. 2000. Financial packaging of IPO firms in China. *Journal of Accounting Research* 38 (1): 103–126.

Allen, F., and G. R. Faulhaber. 1989. Signaling by underpricing in the IPO market. *Journal of Financial Economics* 23 (2): 303–323.

———, J. Qian, and M. Qian. 2005. Law, finance, and economic growth in China. *Journal of Financial Economics* 77 (1): 57–116.

Ball, R., A. Robin, and J. S. Wu. 2003. Incentives versus standards: Properties of accounting income in four East Asian countries. *Journal of Accounting and Economics* 36 (1–3): 235–270.

———, and L. Shivakumar. 2005. Earnings quality in U.K. private firms. *Journal of Accounting and Economics* 39 (1): 83–128.

———, and ———. 2008. Earnings quality at initial public offerings. *Journal of Accounting and Economics* 45 (2–3): 324–349.

Baron, D. P. 1982. A model of the demand for investment banking advising and distribution services for new issues. *The Journal of Finance* 37 (4): 955–976.

Barry, C. B., C. J. Muscarella, J. Peavy, III, and M. R. Vetsuypens. 1990. The role of venture capital in the creation of public companies: Evidence from the going-public process. *Journal of Financial Economics* 27 (2): 447–472.

Beck, T., A. Demirgüç-Kunt, and R. Levine. 2000. A new database on financial development and structure. *The World Bank Economic Review* 14 (September): 597–605.

Benveniste, L. M., and P. A. Spindt. 1989. How investment bankers determine the offer price and allocation of new issues. *Journal of Financial Economics* 24 (2): 343–361.

Bhattacharya, U., H. Daouk, and M. Welker. 2003. The world price of earnings opacity. *The Accounting Review* 78 (3): 641–678.

Biddle, G., and G. Hilary. 2006. Accounting quality and firm-level capital investments. *The Accounting Review* 81 (5): 963–982.

Boulton, T. J., S. B. Smart, and C. J. Zutter. 2010. IPO underpricing and international corporate governance. *Journal of International Business Studies* 41 (February/March): 206–222.

Bradley, D. J., J. W. Cooney, Jr., B. J. Jordan, and A. K. Singh. 2004. Negotiation and the IPO offer price: A comparison of integer vs. non-integer IPOs. *Journal of Financial and Quantitative Analysis* 39 (3): 517–539.

Bushman, R. M., and J. D. Piotroski. 2006. Financial reporting incentives for conservative accounting: The influence of legal and political institutions. *Journal of Accounting and Economics* 42 (1–2): 107–148.

Carter, R. B., and S. Manaster. 1990. Initial public offerings and underwriter reputation. *The Journal of Finance* 45 (4): 1045–1067.

Clarkson, P. M., A. Dontoh, G. Richardson, and S. E. Sefcik. 1992. The voluntary inclusion of earnings forecasts in IPO prospectuses. *Contemporary Accounting Research* 8 (2): 601–616.

Coffee, J. C., Jr. 1999. The future as history: The prospects for global convergence in corporate governance and its implications. *Northwestern University Law Review* 93 (3): 641–708.

——— 2002. Racing towards the top? The impact of cross-listings and stock market competition on international corporate governance. *Columbia Law Review* 102 (7): 1757–1831.

DeFond, M., M. Hung, and R. Trezevant. 2007. Investor protection and the information content of annual earnings announcements: International evidence. *Journal of Accounting and Economics* 43 (1): 37–67.

Degeorge, F., J. Patel, and R. J. Zeckhauser. 1999. Earnings management to exceed thresholds. *The Journal of Business* 72 (1): 1–33.

Doidge, C., G. A. Karolyi, and R. M. Stulz. 2007. Why do countries matter so much for corporate governance? *Journal of Financial Economics* 86 (1): 1–39.

DuCharme, L. L., P. H. Malatesta, and S. E. Sefcik. 2001. Earnings management: IPO valuation and subsequent performance. *Journal of Accounting, Auditing & Finance* 16 (4): 369–396.

Dyck, A., and L. Zingales. 2004. Private benefits of control: An international comparison. *The Journal of Finance* 59 (2): 537–600.

Ellul, A., and M. Pagano. 2006. IPO underpricing and after-market liquidity. *Review of Financial Studies* 19 (2): 381–421.

Fan, Q. 2007. Earnings management and ownership retention for initial public offering firms: Theory and evidence. *The Accounting Review* 82 (1): 27–64.

Friedlan, J. M. 1994. Accounting choices of issuers of initial public offerings. *Contemporary Accounting Research* 11 (1): 1–31.

Gelos, R. G., and S. Wei. 2005. Transparency and international portfolio holdings. *The Journal of Finance* 60 (6): 2987–3020.

Gow, I., G. Ormazabal, and D. J. Taylor. 2010. Correcting for cross-sectional and time-series dependence in accounting research. *The Accounting Review* 85 (2): 483–512.

Grinblatt, M., and C. Y. Hwang. 1989. Signaling and the pricing of new issues. *The Journal of Finance* 44 (2): 393–420.

Hung, M. 2000. Accounting standards and value relevance of financial statements: An international analysis. *Journal of Accounting and Economics* 30 (3): 401–420.

Jagannathan, R., and A. Sherman. 2006. Why do IPO auctions fail? Working Paper No. 12151. Cambridge, MA: NBER.

Jin, L., and S. C. Myers. 2006. R^2 around the world: New theory and new tests. *Journal of Financial Economics* 79 (2): 257–292.

Jog, V., and B. McConomy. 2003. Voluntary disclosure of management earnings forecasts in IPO prospectuses. *Journal of Business Finance & Accounting* 30 (1–2): 125–168.

La Porta, R., F. Lopez-De-Silanes, A. Shleifer, and R. Vishny. 1998. Law and finance. *The Journal of Political Economy* 106 (December): 1113–1155.

———, ———, ———, and ———. 2000. Investor protection and corporate governance. *Journal of Financial Economics* 58 (1–2): 3–27.

Leone, A. J., S. Rock, and M. Willenborg. 2007. Disclosure of intended use of proceeds and underpricing. *Journal of Accounting Research* 45 (March): 111–153.

Leuz, C., D. Nanda, and P. D. Wysocki. 2003. Earnings management and investor protection: An international comparison. *Journal of Financial Economics* 69 (3): 505–527.

Ljungqvist, A. P. 2007. IPO underpricing. In *Handbook of Corporate Finance, Volume 1: Empirical Corporate Finance*, edited by E. Eckbo, 375–422. New York, NY: Elsevier/North Holland.

———, T. Jenkinson, and W. J. Wilhelm, Jr. 2003. Global integration in primary equity markets: The role of U.S. banks and U.S. investors. *Review of Financial Studies* 16 (1): 63–99.

Loughran, T., J. R. Ritter, and K. Rydqvist. 1994. Initial public offerings: International insights. *Pacific Basin Journal* 2 (2–3): 165–199.

———, and ———. 2004. Why has IPO underpricing changed over time? *Financial Management* 33 (Autumn): 5–37.

Megginson, W. L., and K. A. Weiss. 1991. Venture capitalist certification in initial public offerings. *The Journal of Finance* 46 (3): 879–903.

Moeller, S. B., and F. P. Schlingemann. 2005. Global diversification and bidder gains: A comparison between cross-border and domestic acquisitions. *Journal of Banking & Finance* 29 (March): 533–564.

Prezas, A. P., M. Tarimcilar, and G. K. Vasudevan. 2000. The pricing of equity carve-outs. *Financial Review* 35 (4): 123–138.

Ritter, J. R. 1984. The hot issue market of 1980. *The Journal of Business* 57 (2): 215–240.

——— 1987. The costs of going public. *Journal of Financial Economics* 19 (2): 269–281.

——— 1991. The long-run performance of initial public offerings. *The Journal of Finance* 46 (1): 3–27.

Rock, K. 1986. Why new issues are underpriced. *Journal of Financial Economics* 15 (1–2): 187–212.

Ruud, J. S. 1993. Underwriter price support and the IPO underpricing puzzle. *Journal of Financial Economics* 34 (2): 135–151.

Schipper, K., and A. Smith. 1986. A comparison of equity carve-outs and seasoned equity offerings: Share price effects and corporate restructuring. *Journal of Financial Economics* 15 (1–2): 153–186.

Schrand, C., and R. E. Verrecchia. 2005. Information disclosure and adverse selection explanations for IPO underpricing. Working paper, University of Pennsylvania.

Sherman, A. E. 2005. Global trends in IPO methods: Book building versus auctions with endogenous entry. *Journal of Financial Economics* 78 (3): 615–649.

Teoh, S. H., I. Welch, and T. J. Wong. 1998a. Earnings management and the long-run market performance of initial public offerings. *The Journal of Finance* 53 (6): 1935–1974.

———, T. J. Wong, and G. R. Rao. 1998b. Are accruals during initial public offerings opportunistic? *Review of Accounting Studies* 3 (1–2): 175–208.

———, and ———. 2002. Why new issues and high-accrual firms underperform: The role of analysts' credulity. *Review of Financial Studies* 15 (3): 869–900.

Venkataraman, R., J. P. Weber, and M. Willenborg. 2008. Litigation risk, audit quality, and audit fees: Evidence from initial public offerings. *The Accounting Review* 83 (5): 1315–1345.

Welch, I. 1989. Seasoned offerings, imitation costs, and the underpricing of initial public offerings. *The Journal of Finance* 44 (2): 421–449.

Willenborg, M., and J. C. McKeown. 2001. Going-concern initial public offerings. *Journal of Accounting and Economics* 30 (December): 279–313.

THE ACCOUNTING REVIEW
Vol. 86, No. 2
2011
pp. 507–539

American Accounting Association
DOI: 10.2308/accr.00000017

Are Investors Confused by Restatements after Sarbanes-Oxley?

Jeffrey J. Burks
University of Notre Dame

ABSTRACT: Regulators have expressed concern that investors are confused by the large number and questionable materiality of accounting restatements since passage of the Sarbanes-Oxley Act (SOX). This study looks for evidence of investor confusion by examining stock returns and trading volume. I find that the initial price reaction to restatement announcements becomes significantly less negative after SOX, even after controlling for the less egregious nature of post-SOX restatements. To assess whether the less negative reaction represents under-reaction, I test whether stock prices drift negatively over the months and years after the initial reaction. I find no evidence of statistically negative drifts unique to the post-SOX period. In fact, I find that post-SOX drifts are statistically less negative than pre-SOX drifts, suggesting that price efficiency actually improves after SOX. Finally, I find no evidence that post-SOX restatements have higher trading volume after controlling for contemporaneous returns, suggesting no increase in disagreement among investors about the restatements. Thus, the findings provide little evidence that investors are confused by post-SOX restatements.

Keywords: *accounting changes; accounting errors; restatements; Sarbanes-Oxley Act; regulation.*

Data Availability: *Data are available from public sources.*

I. INTRODUCTION

The sharp increase in the number of accounting restatements since passage of the Sarbanes-Oxley Act (SOX, U.S. House of Representatives 2002) has received considerable regulatory attention. In an op-ed piece for the *Financial Times*, United States Treasury Secretary Henry Paulson expressed concern that many of these restatements could needlessly alarm

I am thankful for the helpful comments of Steven Kachelmeier and two anonymous reviewers, participants at the 2009 FARS Midyear Meeting, the 2009 AAA Annual Meeting, Peter Demerjian (the FARS discussant), Brian Hogan (the AAA discussant), and anonymous reviewers for the meetings. I also thank workshop participants at the University of Nebraska and University of Notre Dame, and Robert Battalio, Peter Easton, Geoffrey Friesen, Bruce Johnson, Jim Seida, and Alice Heerse of Standard & Poor's. Thanks to Sarah Anderson for capable research assistance.

Editor's note: Accepted by Steven Kachelmeier.

Submitted: March 2009
Accepted: July 2010
Published Online: March 2011

investors, leading to "confusion" over the quality of the financial reports.[1] Similarly, the Securities and Exchange Commission's Advisory Committee on Improvements to Financial Reporting (CIFR) fears that restatements "may create confusion that reduces the efficiency of investor analysis" (CIFR 2008, 78).

Regulators' concerns have a theoretical basis. Restatements likely increase processing costs for both institutional and individual investors, and several studies theorize that investors have difficulty impounding costly information into asset prices (Grossman and Stiglitz 1980; Bloomfield 2002; Hirshleifer and Teoh 2003). Due in part to these concerns, CIFR recommends reducing the scope of violations that necessitate a restatement. However, these recommendations are opposed by some investor groups that believe investors benefit from the restatements' transparent portrayal of prior period effects (Johnson 2008).

This study looks for signs of investor confusion in the price and volume reactions to post-SOX restatements, which help regulators assess whether investor confusion is a valid rationale for curtailing restatements. The pricing tests examine whether investors systematically over- or under-react to restatements. Although Paulson's comments suggest a fear that investor reaction to the increased number of restatements will be overly negative, investors could also respond by ignoring or under-reacting to restatements due to the additional processing costs. CIFR Chairman Robert C. Pozen expresses uncertainty over whether investors would over- or under-react: "A restatement imposes obvious costs on a preparer, but it also imposes two sorts of costs on investors. One is confusion: If a restatement used to mean to investors 'this is a really big problem,' and you get a lot of restatements that aren't big problems, then we haven't sorted that out" (Cummings 2008, 1).[2] Pozen also suggests that, with so many restatements to track, investors may fail to distinguish the egregious restatements from those that are more innocuous (Heffes 2007).[3] "Confused" investors who fail to discern differences in egregiousness across restatements would likely under-react to the more egregious restatements and over-react to the less egregious restatements.

I test for over- or under-reaction primarily by examining long-run stock price drifts following restatements. However, I first examine the initial stock price reactions to restatements, benchmarking the initial reactions in the post-SOX period against those in pre-SOX period, when restatements were less frequent and firms faced less pressure to restate for errors of questionable materiality. Prior studies note that price reactions to restatement announcements become less negative after SOX (U.S. Government Accountability Office [GAO] 2006; Hranaiova and Byers 2007; Scholz 2008), which is consistent with investors under-reacting in the post-SOX period. However, the less negative reactions may be warranted if post-SOX restatements are less egregious, as regulators claim. I find that, after controlling for the less egregiousness nature of post-SOX restatements, post-SOX reactions continue to be less negative.

If less negative initial reactions are truly under-reactions, then stock prices should continue to drift downward as prices converge to fundamental value. Contrary to claims of higher investor confusion after SOX, I find that post-SOX restatements have significantly less drift than do pre-SOX restatements. I also test for drifts in subsamples of restatements, in case the full sample

---

[1] Paulson stated, "Restatements pose significant costs on our capital markets. They have the potential to confuse investors and erode public confidence in financial reporting. Some of these restatements might not be material to investors, and others may simply reflect new accounting standards interpretations" (Paulson 2007, 15).

[2] The second cost to investors Pozen mentions is the delay in earnings announcements and other disclosures while firms undergo the process of restating financial statements. Badertscher and Burks (2010) examine this issue.

[3] Pozen stated, "I think the evidence is that many restatements seem to have very little market effect. On the other hand, there are some restatements with significant market impacts, so we should be helping preparers and investors sort out which ones are which" (Heffes 2007, 17).

exhibits no drift simply because investors under-react to more egregious restatements and over-react to less egregious restatements. Consistent with the full sample result, the subsample analysis provides no evidence of drift unique to the post-SOX period.

The pricing tests may not detect investor confusion because prices aggregate investor beliefs. Prices can be efficient even in a market containing confused investors if overly optimistic investors trade with overly pessimistic investors, or if the market includes rational arbitrageurs with enough power to correct the influence of the confused investors. In either case, the presence of confused investors would lead to trading volume that is independent of price movements (Kim and Verrecchia 1997). Thus, as a final test for investor confusion, I test for an increase in trading volume after controlling for price movements around post-SOX restatement announcements. I do not find a statistically discernable trading volume increase for post-SOX restatements. In summary, the results of the pricing and volume tests do not support the notion that the increase in restatements after SOX produced investor confusion.

This study contributes to the debate between regulators and investor groups over proposed reforms that would curtail the use of restatements for accounting errors. Regulators' concerns about investor confusion have fostered momentum for the reforms. However, this study's finding that investors price restatements more efficiently after SOX than before suggests that investor confusion is not a valid rationale for curtailing restatements.

The study also informs a broader standard-setting question about investors' capacity to process restatements for accounting changes other than error corrections. Fears about investor confusion have caused U.S. standard-setters to prohibit restatements for many types of accounting changes since 1971 (Accounting Principles Board [APB] Opinion No. 20, APB 1971, para. 14b). In 2005 the Financial Accounting Standards Board partially reversed course and began requiring firms to restate for changes in accounting principle (Statement of Financial Accounting Standards [SFAS] No. 154, FASB 2005). Some academics have advocated further expanding the use of restatements to changes in accounting estimates (Lundholm 1999; Lev 2003). My study suggests that investors were able to adapt to a dramatic increase in restatements for accounting errors, leaving open the possibility that investors could adapt to an environment with even more types of restatements.

II. BACKGROUND

Controversy Over the Increase in Restatements after SOX

Studies by the GAO of the period 1997 to 2005 show a marked increase in restatements beginning in 2002, the year SOX was passed (GAO-06-678 [GAO 2006], GAO-06-1053R [GAO 2007]). This increase is depicted in Figure 1 and is discussed further in Section IV.[4] Since the increase in restatements begins in 2002, I refer to it as a "post-SOX" phenomenon, although SOX may not be the sole cause. Both SOX and the related high-profile accounting scandals (e.g., Enron and WorldCom) likely prompted managers and auditors to become more vigilant in detecting and correcting accounting errors, leading to more restatements.

Regulators have several concerns about the increase in restatements, including the potential for investor confusion, the burden on preparers, and reporting delays involved with the restatement process (CIFR 2008). This study complements other academic work addressing these concerns (Wright et al. 2008; Badertscher and Burks 2010) by testing for evidence of investor confusion, a particular point of contention between regulators and investor groups.

---

[4] The GAO sample ends in 2005. Tracking by research firm Glass, Lewis & Co., LLC and by Audit Analytics, a research arm of the Ives Group, Inc., shows that restatements peaked in 2006, but remained high in 2007 and 2008, dwarfing the restatement counts of years in the late 1990s by factors of 5 to 10 (Taub 2009).

FIGURE 1
Restatement Counts by Year

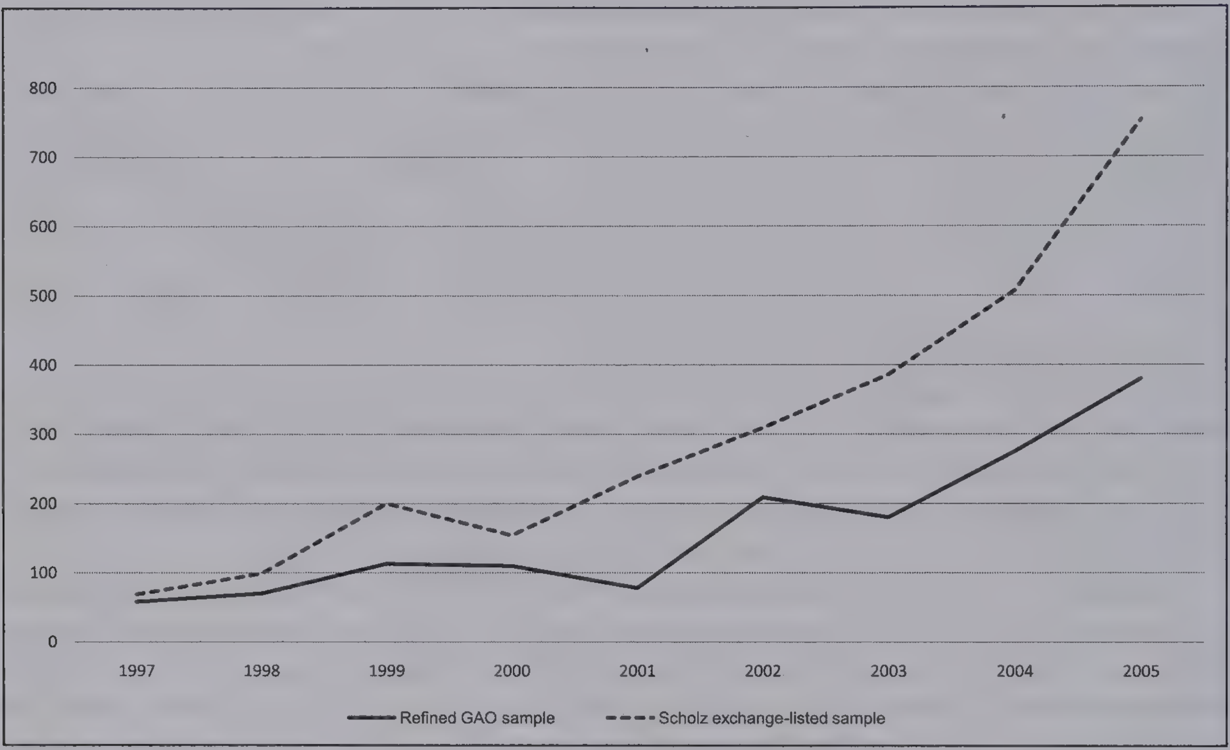


Figure 1 shows sample size by year for the GAO sample after the refinements described in Table 1 are applied. For comparison, sample counts for exchange-listed companies from Scholz (2008), Figure 1 are shown.

Policymakers' opposition to revisions of prior-period financial statements dates back to APB Opinion No. 20 in 1971 (Lev et al. 2008), and stems from concern about how investors perceive and process restatements: "Restating financial statements of prior periods may dilute public confidence in financial statements and may confuse those who use them" (APB Opinion No. 20, para. 14b). Thus, the APB prohibited restatements for changes in accounting principle and estimates, but did require restatements for accounting errors.

As the number of restatements ballooned after 2002, concerns about investor confusion resurfaced and played a role in controversial proposals to reform restatement practices. After Secretary Paulson's op-ed piece in the *Financial Times*, the Treasury Department commissioned a study examining the increase in restatements and capital market effects (Scholz 2008). The SEC formed the CIFR, which recommended ways to reduce restatements, proposing that more errors be classified as immaterial so restatements would not be necessary (SFAS No. 154). Under the CIFR's proposals, even if the cumulative impact of an error is large, the error would be classified as immaterial if it relates to business activities that do not drive firm value or risk, or if the error is immaterial to each prior period viewed in isolation (CIFR 2008, 81–83). Some observers, including a former SEC Chief Accountant, claim that the SEC has informally adopted these or other restatement-reducing policies (Rummell 2008; Stuart 2010).

Many investor groups oppose curtailing restatements, claiming that investors are capable of assessing their value relevance. At a public forum hosted by CIFR, an analyst for the Capital

Group Cos. stated that "investors want to be their own decision-makers of which errors are unimportant in their investment theses" (Johnson 2008, 1). In a comment letter to the CIFR, the Consumer Federation of America (CFA) stated that attempting to eliminate so-called unnecessary restatements "poses risks to investors that far exceed any potential benefits" (CFA 2008, 3). The Chartered Financial Analyst Institute (CFA Institute) state that "it is better to err on the side of providing too much disclosure rather than too little. We note that investors use a wide variety of investment approaches and models, so attempting to determine if certain changes might alter certain approaches could be problematic" (CFA Institute 2008, 19).

Despite these assertions, regulators' concerns about investors' ability to value restatements may have merit. To efficiently price restatements, investors must convert the revised financial statements to usable form. D'Souza et al. (2010) emphasize that many institutional investors rely on data aggregators like Standard & Poor's (S&P) to organize financial information into a standardized machine-readable form. However, as explained to me by an S&P manager, S&P's collection procedures prioritize current filings ahead of amended filings. Amended filings made during peak times of the year often take forty to eighty days to reach clients.

Consequently, investors who want immediate access to restated financial statements would need to set up their own processes for monitoring amended filings, inputting data, and controlling quality control, endeavors that become more costly as the number of restatements increases. The S&P manager reports that monitoring amended filings is difficult because they are not scheduled like earnings announcements, and are made for a host of reasons other than accounting restatements, such as missing signatures or unsatisfactory qualitative disclosures.[5] The restatement footnote within the amended filing often does not provide restated versus original comparisons for all periods affected. To identify changes in longer historical trends, the investor must check each figure in the five-year summary table of "selected financial data" (item 6 of the 10-K) to determine whether it has been restated, a process the S&P manager describes as "cumbersome." If a data aggregator like S&P finds these steps difficult, then many institutional investors likely find them difficult as well.

Mispricing could result even if arbitrageurs are willing to bear the collection costs. Black (1986) and Shleifer and Vishny (1997) discuss general limits to risky arbitrage, and several studies specifically focus on the limits related to costly information. Hirshleifer and Teoh (2003) examine a setting in which investors must process vast amounts of available information, and show that investors with "limited attention" create mispricing that fully attentive investors do not arbitrage away because of risk aversion. Grossman and Stiglitz (1980) argue that there is an equilibrium degree of mispricing in markets because arbitrageurs require trading gains as compensation for costly information searches. They show that the degree of mispricing increases in the cost of information. Bloomfield (2002) also bases his "incomplete revelation hypothesis" on the notion that arbitrageurs must be compensated for costly information search; the hypothesis holds that investors can misprice even widely publicized events like restatements because of "the costs of identifying, collecting, compiling, printing and processing data, or hiring others to do so" (Bloomfield 2002, 236). In light of the apparent increase in information costs that accompanies the large number of post-SOX restatements, my study tests whether price efficiency declines after SOX.

[5] In August 2004, the SEC made accounting restatements easier to identify by requiring announcement in a Form 8-K when management concludes that prior financial statements should not be relied upon. However, firms still do not file 8-Ks for nearly one-third of restatements even after passage of the rule (Myers et al. 2010). Even when an 8-K is filed, the firm may not file amended financial statements until several weeks or months later, requiring investors to keep watch.

Prior Studies on Market Reaction to Restatements

One stream of prior research shows that the average market reaction to restatement announcements becomes less negative after SOX (GAO 2006; Hranaiova and Byers 2007; Scholz 2008). A second stream of research finds that restatements become, on average, less egregious after SOX; post-SOX restatements commonly involve lower dollar amounts, unintentional errors, and non-core accounts (Burks 2010; Hennes et al. 2008; Plumlee and Yohn 2010; Scholz 2008). Several studies document that these restatement characteristics affect market reaction (e.g., Palmrose et al. 2004a; Gordon et al. 2008; Hennes et al. 2008; Files et al. 2009). Thus, I merge the two streams of literature to determine whether the less egregious nature of post-SOX restatements can completely explain the less negative market reactions. Two concurrent studies also examine reasons for the decline in market reaction (Hirschey and Smith 2009; Huang 2010). Like my study, both find that changes in restatement characteristics do not fully explain the less negative post-SOX reactions. A unique contribution of my study is that it tests whether mispricing, a chief concern of regulators, is the reason that changes in restatement characteristics fail to explain the decline in market reaction.

Another unique feature of my study is that it tracks the date when firms disclose the restatement's earnings impact. Firms often do not disclose an error's full impact at the initial announcement of a pending restatement; the process of quantifying the errors can take several weeks or months. Prior studies focus on the price reaction during a relatively short window around the initial announcement when investors often do not have complete information about the restatement. I examine the price reaction using a short event window as in prior studies, but also use an "episode" window that spans from the initial restatement announcement to the release date of the earnings impact. I also incorporate the release date of the earnings impact into the long-run price drift tests, to distinguish drifts due to slow release of information from drifts due to investors mispricing released information.

III. RESEARCH DESIGN

Announcement Returns

I use the following model to assess whether restatement market reactions become less negative after SOX simply because restatements become less egregious:

$$Return_i = \alpha + \beta_1 POSTSOX_i + \beta_2 RSTMT\ CONTROLS_i + \beta_3 GENERAL\ CONTROLS_i + \varepsilon_i. \quad (1)$$

$Return_i$ is the size-adjusted, buy-and-hold stock return of firm i measured over one of two windows: the three-day window $(-1, +1)$ around the initial restatement announcement (*RETURN_ANNOUNCE*), or a window beginning one day before the initial restatement announcement and ending one day after the earnings impact announcement (*RETURN_EPISODE*). The earnings impact announcement is when a firm reveals the restatement's impact on past earnings, either cumulatively or by period. The announcement must be definitive, not an estimate, and can come in a press release before the firm actually files amended financial statements with the SEC.

The variable of interest is *POSTSOX*, an indicator that captures whether the restatement is announced in the month SOX was passed (July 2002) or after. Although prior studies document less negative announcement returns to post-SOX restatements, I do not predict the *POSTSOX* coefficient sign because announcement returns could be higher, lower, or the same after controlling for the post-SOX decline in egregiousness.

I briefly describe two sets of control variables in this section; the Appendix includes detailed descriptions of the variables. The first set of controls relates to restatement characteristics. I include the magnitude of the restatement's earnings impact scaled by total assets (*MAG*), an indicator variable for nonnegative earnings impacts (*POSITIVE_MAG*), and an interaction be-

tween them. *MAG* and *POSITIVE_MAG* are expected to be positively related to announcement returns. I expect a negative sign on the interaction term's coefficient because earnings-increasing restatements are not necessarily good news and so may not garner positive returns, thus attenuating *MAG*'s effect. I also include a term that multiplies *MAG* by its absolute value ($MAG * |MAG|$) in case investors do not react as strongly per dollar to restatements with extreme earnings impacts (Lipe et al. 1998).

I expect the following restatement characteristics to positively affect returns: whether the restatement is prompted by management as opposed to the auditor or SEC (*MGT_PROMPTS*) and whether the restatement involves quarters of the current fiscal year only (*QUARTERLY_ONLY*). I expect the following restatement characteristics to negatively affect returns: whether the definitive earnings impact is unknown at the time of the initial announcement (*NO_INITIAL_MAG*); whether the firm takes more than 30 days to disclose the earnings impact after the initial announcement (*LONG_MAG_DELAY*); whether the restatement results from fraudulent actions at the corporate level (*FRAUD*) or lower (*FRAUD_LOWLEVEL*); whether the restatement affects components of pre-tax operating income other than non-cash items like depreciation (*CORE*); and whether the restatement involves multiple errors (*MULTIPLE*). To form *MULTIPLE*, I place each restatement into one of 11 mutually exclusive categories based on the items restated. *MULTIPLE* equals 1 when the restatement spans categories or involves three or more errors in the same category.

The second set of controls relates to general factors that potentially affect the market reaction to restatements. To control for pre-announcement information leakage or industry-spillover effects (Gleason et al. 2008), I use size-adjusted stock returns in the 180 days preceding the announcement (*RETURN_PRIOR*), expecting a negative coefficient. I use the natural log of market capitalization (*FIRMSIZE*) to control for reputational effects that could mitigate negative reactions to restatements (Gordon et al. 2008). Following Scholz (2008), I include the mean value of the Chicago Board Options Exchange's volatility index over the return window (*VOLATILITY_INDEX*) because reactions to bad news may be more pronounced when market volatility is high. Finally, I control for surprises associated with any earnings announcements in the return window (*ESURPRISE* or *ESURPRISE_EPISODE*), and whether the earnings announcements meet or beat the consensus estimate (*MBE*). *MBE* is not included in the episode return regression because the episode often contains more than one earnings announcement.

Long-Run Price Drifts

I test the efficiency of investor reaction to restatements by investigating post-announcement price drift.[6] I measure drifts using Lyon et al.'s (1999) method for computing abnormal buy-and-hold returns. First, I restrict the analysis to ordinary common shares (CRSP share codes 10 and 11). Then I form benchmark portfolios by partitioning all CRSP/Compustat NYSE firms into size-based deciles using market value of common equity at the beginning of the calendar year.[7] I then add Nasdaq and AMEX firms to the deciles (using the NYSE breakpoints), and partition the lowest decile into quintiles, which creates a total of 14 size portfolios. I then divide each size portfolio into book-to-market quintiles, creating 70 benchmark portfolios.

To compute the return for a benchmark portfolio, I compute the buy-and-hold return over the horizon for each firm in the portfolio and then average across firms. If the firm stops trading before the horizon ends, then I incorporate the delisting return as in Beaver et al. (2007) and use the

---

6 Price drifts could result from compensation for unidentified risk factors rather than mispricing. However, to the extent any unidentified risk factor operates similarly in the pre- and post-SOX periods, such a factor could not explain the *differences* in drift (the focus of this study).

7 Lyon et al. (1999) compute market value of equity in July of each year. I use the beginning of the year instead because many restatements are announced in the first quarter, so using July may create a look-ahead bias.

portfolio-level return for the remaining days.[8] I do not rebalance or allow new firms to enter the benchmark portfolios within the horizon because this induces bias (Lyon et al. 1999). I then subtract the benchmark portfolio's buy-and-hold return from the restatement firm's buy-and-hold return to compute the abnormal buy-and-hold return.

Lyon et al. (1999) find that skewness in buy-and-hold returns affects statistics that rely on normality, and recommend evaluating significance using an empirical distribution of 1,000 "pseudo-portfolios." Each pseudo-portfolio consists of a set of control firms that matches the size and book-to-market profile of the set of restatement firms. To form a single pseudo-portfolio of control firms, I match each restatement firm randomly with replacement to a non-restatement firm in the same size/book-to-market benchmark portfolio. Then I compute an abnormal return for each control firm in the pseudo-portfolio as described above, and find the median abnormal return of the pseudo-portfolio. This process is repeated 1,000 times to form an empirical distribution of median abnormal returns. I then assess the statistical significance of the restatement firms' median abnormal return using its ranking in the empirical distribution.[9] I focus on medians to reduce the influence of outliers; the study seeks to test whether investors typically under-react or over-react to restatements, not whether a trading strategy to exploit the mispricing is executable.

Abnormal Volume

Even if prices are efficient, confusion could still exist among investors because prices reflect aggregate, not individual, beliefs. In the aggregate, the beliefs of rational investors may dominate those of confused investors, or the overly positive beliefs of some confused investors may offset the overly negative beliefs of other confused investors. Kim and Verrecchia (1997) model how these types of disagreements across investors cause higher levels of trading volume for a given level of contemporaneous returns. The idea is that disagreements over underlying firm value generate trading, but the lack of consensus hinders movements in price (Beaver 1968).[10] If confusion over post-SOX restatements reduces consensus across investors, then post-SOX restatements should have higher trading volume after controlling for contemporaneous returns. I test for this effect using the following model:

$$VOLUME_i = \alpha + \beta_1 POSTSOX_i + \beta_2 RSTMT\ CONTROLS_i + \beta_3 GENERAL\ CONTROLS_i + \varepsilon_i. \tag{2}$$

To adjust the dependent variable for market-wide volume and the firm's normal volume, I follow the method of Garfinkel and Sokobin (2006): first scale the firm's daily volume by shares outstanding to compute "turnover," or the percentage of shares outstanding traded that day; compute a market-wide turnover measure using NYSE and AMEX common stocks, and subtract this from the firm's turnover; and then compute average market-adjusted turnover (TO) over the three-day window around the restatement announcement:

$$TO_{restate,i} = \left\{ \sum_{t=-1}^{1} \left[\left(\frac{Vol_{i,t}}{Shs_{i,t}}\right)_{firm} - \left(\frac{Vol_t}{Shs_t}\right)_{mkt} \right] \right\} \Big/ 3 \tag{3}$$

where $Vol_{i,t}$ is firm i's volume on day t, and $Shs_{i,t}$ is firm i's shares outstanding on day t. Finally, to adjust for the firm's normal level of TO, compute the average TO over trading days (−250,

[8] If the delisting return is missing, I use a replacement value based on the mean for that delisting code from Beaver et al. (2007, Table 3).

[9] Similar pseudo-portfolio methods are used by Ikenberry et al. (1995), Lee (1997), and Byun and Rozeff (2003).

[10] Other studies that infer disagreement from higher trading volume after controlling for returns include Bamber and Cheon (1995), Kandel and Pearson (1995), Bamber et al. (1997), and Bailey et al. (2003).

−31) relative to the restatement announcement (TO_{pre}), and subtract from the announcement window *TO*:

$$VOLUME_i = TO_{restate,i} - TO_{pre,i}. \tag{4}$$

The explanatory variable of interest in Equation (2) is *POSTSOX*; a positive coefficient sign is consistent with post-SOX restatements generating more disagreement among investors. The main control variable is the absolute value of the abnormal return in the three-day restatement announcement window (*AB_RETURN*). I interact *AB_RETURN* with an indicator variable for positive abnormal returns (*POSITIVE_RETURN*) to allow for different volume reactions to good and bad news (Karpoff 1987).[11] Following Bailey et al. (2003), I also interact *AB_RETURN* with *POSTSOX* in case the relation between volume and returns changes after SOX.

As additional controls for news, I include the restatement characteristics from the model of announcement returns, using absolute values for earnings magnitude (*AB_MAG*), earnings surprise (*AB_ESURPRISE*), and returns leading up to the announcement (*AB_RETURN_PRIOR*). *AB_RETURN_PRIOR* is expected to negatively affect *VOLUME* because large unsigned returns before the restatement suggest news leakage that reduces the news in the restatement announcement. Because the revelation of erroneous accounting is fundamentally bad news, I expect variables that worsen the news associated with a restatement to positively affect volume (*FRAUD*, *LONG_MAG_DELAY*, etc.), and variables that lessen the bad news to negatively affect volume (*POSITIVE_MAG*, *MGT_PROMPTS*, and *QUARTERLY_ONLY*). Finally, I expect the *VOLATILITY_INDEX* to be positively related to volume because restatement news could generate more disagreement among investors during times of high uncertainty.

IV. SAMPLE SELECTION

To form the sample, I use two reports from the GAO (GAO-03-138 [GAO 2002] and GAO-06-678 [GAO 2006]) that identify restatements from 1997 to September 2005. The two reports use consistent methods for identifying restatements across the pre- and post-SOX periods.[12] The reports focus on restatements of exchange-traded firms, which improves the chances that firm data are available on CRSP, Compustat, and I/B/E/S.

Before hand-collecting information about each restatement, I eliminate 323 of the 2,309 restatements because they lack basic data on CRSP and Compustat in the year of or year before the restatement. I eliminate another 513 restatements in the process of hand-collecting, leaving 1,473 restatements. The most common reasons for elimination are that the error relates to an earnings release for the current period rather than a prior period 10-K or 10-Q (114), or the GAO captures more than one announcement for the same restatement (107). Table 1 lists other reasons for elimination.

The GAO reports the announcement date of each restatement. However, I find earlier disclosures mentioning the possibility of restatement for 250 of the 1,473 restatements, with a mean (median) difference of 47 (22) days. I identify earlier disclosures because the GAO requires companies to imply that a restatement is at least "likely" (GAO-06-678, 56), whereas I look for the

[11] Based on prior studies, Karpoff (1987) concludes that volume is higher around good news events than bad news events. I do not predict a sign for my *AB_RETURN * POSITIVE_RETURN* interaction because the bad-news nature of restatements means that my sample is fundamentally different from studies Karpoff (1987) reviews, which examine the general relation between volume and price changes.

[12] Although the GAO issued the reports several years apart, in the second report the GAO states, "To determine the number of and reasons for restatements since 2002, we employed substantially the same methodology used in our prior report, in which we analyzed the period from January 1997 through June 2002" (GAO 2006, 52).

TABLE 1

Sample Selection

Restatements Identified by the GAO from 1997 to September 2005	**2,309**
1. Missing basic Compustat and CRSP data in year of or year before restatement	−323
2. Firm is amending an earnings release rather than a prior form 10-K or 10-Q	−114
3. Subsequent announcements related to the same restatement	−107
4. Adopting SAB No. 101	−72
5. Adopting a new standard	−65
6. Firm never files restated financials because of bankruptcy, acquisition, etc.	−33
7. Restatement impact is not released in US dollars	−25
8. Changing from one within-GAAP method to another	−24
9. Firm is not an SEC filer	−16
10. Firm decides not to restate after the initial announcement	−16
11. Other	−41
Base Sample	1,473

Observations vary by table depending on data availability. I exclude restatements due to adoption of SEC Staff Accounting Bulletin No. 101 (step 4 above) because of controversy over whether the bulletin represented new GAAP or was simply a reiteration of existing GAAP (Altamuro et al. 2005).

first mention of possible accounting errors. Using the earliest date mentioned allows me to more accurately capture the market reaction to the entire restatement episode, from the first mention of accounting problems to the revelation of the earnings impact.

Figure 1 plots the number of restatements by year in my sample and includes a comparison plot of the Scholz (2008) sample. Two differences are apparent. First, the GAO sample has fewer restatements each year. The GAO has fewer restatements from 1997 to 2000 likely because Scholz (2008) supplements the GAO list for these years with her own Lexis-Nexis searches (Scholz 2008, 8). Beginning in 2001, Scholz (2008) switches to the Audit Analytics database, which uses procedures that are more comprehensive than the GAO's to identify restatements. The large difference in 2005 likely arises because the GAO ends its search in September of that year but Audit Analytics covers the entire year. The observations I eliminated in Table 1 may also account for some of the differences.

The second major difference between the two samples is that the increase in restatement activity begins in 2002 for the GAO sample and 2001 for the Scholz (2008) sample. Scholz (2008) reports that the increase in her sample may be caused by the switch to Audit Analytics in 2001, as Audit Analytics uses more thorough search technologies and often includes multiple announcements for the same restatement (Scholz 2008, 8, 11).[13]

V. RESULTS

Descriptive Statistics

Data requirements for the regressions further reduce the sample of restatements from 1,473 to 1,226.[14] The pre-SOX (post-SOX) period contains 407 (819) restatements. Table 2 presents pre- and post-SOX comparisons of the variables. To reduce the influence of outliers, I winsorize the

[13] Scholz (2008) attempts to remove duplicate announcements by eliminating subsequent observations for the same firm over the next 90 days.

[14] Missing I/B/E/S information causes most of the sample reduction. Regression results are similar when using the larger sample without the I/B/E/S-related controls.

TABLE 2
Descriptive Statistics

	Pre-SOX (n = 407)			Post-SOX (n = 819)			Difference	
	Mean	Median	Std. Error	Mean	Median	Std. Error	Mean	Median
Dependent Variables								
RETURN_ANNOUNCE	−0.102**	−0.060**	0.167	−0.033**	−0.015**	0.106	0.069**	0.046**
RETURN_EPISODE	−0.147**	−0.071**	0.249	−0.043**	−0.015**	0.171	0.103**	0.056**
VOLUME	0.018**	0.002**	0.052	0.006**	0.001**	0.022	−0.012**	−0.001*
Variable of Interest								
POSTSOX	0.000	0.000	0.000	1.000	1.000	0.000	NA	NA
Control Variables								
Restatement Characteristics								
MAG	−0.041**	−0.011	0.109	−0.023**	−0.004	0.070	0.018**	0.007**
POSITIVE_MAG	0.214	0.000	0.410	0.275	0.000	0.447	0.061*	0.000
NO_INITIAL_MAG	0.568	1.000	0.496	0.637	1.000	0.481	0.070*	0.000
LONG_MAG_DELAY	0.359	0.000	0.480	0.385	0.000	0.487	0.026	0.000
FRAUD	0.349	0.000	0.477	0.214	0.000	0.410	−0.135**	0.000
FRAUD_LOWLEVEL	0.076	0.000	0.266	0.040	0.000	0.197	−0.036**	0.000
MGT_PROMPTS	0.472	0.000	0.500	0.714	1.000	0.452	0.243**	1.000
CORE	0.398	0.000	0.490	0.233	0.000	0.423	−0.165**	0.000
MULTIPLE	0.221	0.000	0.416	0.248	0.000	0.432	0.027	0.000
QUARTERLY_ONLY	0.400	0.000	0.491	0.168	0.000	0.375	−0.232**	0.000
POSTIVE_RETURN	0.268	0.000	0.443	0.353	0.000	0.478	0.085**	0.000
General Controls								
RETURN _PRIOR	−0.184**	−0.239**	0.424	−0.021	−0.023**	0.330	0.163**	0.216**
FIRMSIZE	12.262	12.054	2.081	12.960	13.003	1.870	0.699**	0.949**
VOLATILITY_INDEX	23.982	23.200	4.015	18.336	15.160	7.688	−5.646**	−8.040**
ESURPRISE	−0.002**	0.000	0.013	−0.001**	0.000	0.010	0.001	0.000
ESURPRISE_EPISODE	−0.019**	0.000	0.070	−0.009**	0.000	0.048	0.011*	0.000
MBE	0.128	0.000	0.334	0.188	0.000	0.391	0.060**	0.000

(continued on next page)

TABLE 2 (continued)

*, ** Denote the value is significantly different from 0 at the 5 percent and 1 percent levels, respectively (two-tailed).

t-tests (Wilcoxon signed-rank tests) are used for assessing means (medians). For continuous variables, difference-in-means t-statistics are computed assuming unequal variances, and differences in medians are assessed using Wilcoxon two-sample tests. For dichotomous variables, differences are assessed using Chi-squared tests. Statistical significance is reported only when the null hypothesis is 0. *ESURPRISE_EPISODE* has only 352 pre-SOX and 716 post-SOX observations due to missing data on I/B/E/S.

See the Appendix for descriptions of variables.

following variables at the 1st and 99th percentiles: *RETURN_ANNOUNCE*, *RETURN_EPISODE*, *MAG*, *RETURN_PRIOR*, *ESURPRISE*, and *ESURPRISE_EPISODE*.

Mean restatement announcement returns (*RETURN_ANNOUNCE*) are significantly negative in both periods, but are much less negative in the post-SOX period (−10.2 percent pre-SOX versus −3.3 percent post-SOX), consistent with prior research (GAO 2006; Scholz 2008). Mean returns measured over the entire episode (*RETURN_EPISODE*) are even more negative, and the difference in returns between the two periods is even larger (−14.7 percent pre-SOX versus −4.3 percent post-SOX). The same trends hold for median returns. To understand when the decline in market reaction begins, Figure 2 plots median *RETURN_EPISODE* by year. The decline in reaction roughly coincides with the increase in the number of restatements that began in 2002; median returns for each year from 2002 to 2005 are noticeably less negative than those from 1997 to 2001 (depicted with gray bars in the figure). Among just the non-fraudulent restatements, the drop in

FIGURE 2
Median Episode Returns by Year (*RETURN_EPISODE*)

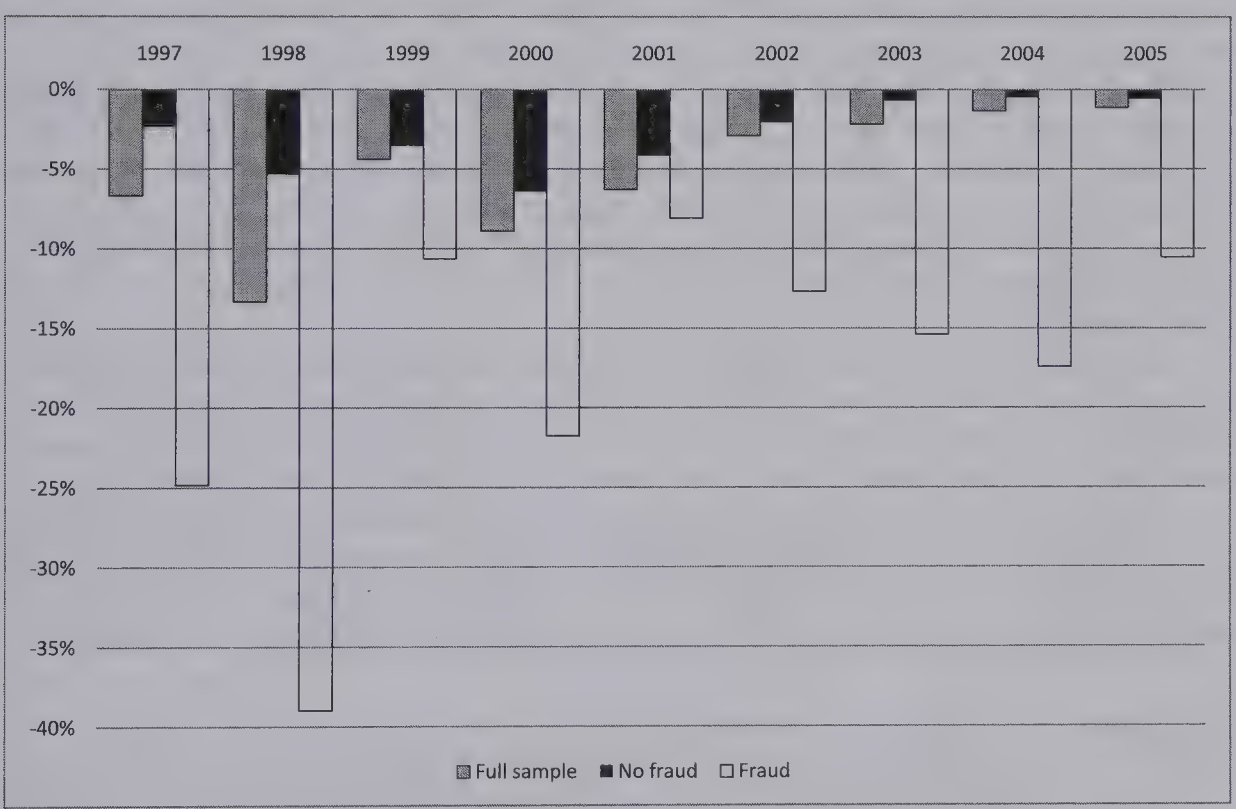


Figure 2 shows the median size-adjusted returns of restatement firms beginning one day before the restatement announcement and ending one day after the earnings impact announcement.

median returns occurs in 2002 or 2003 (depicted with black bars). No trend is evident in the returns of fraudulent restatements (depicted with white bars).[15]

Mean and median abnormal *VOLUME* in the three-day window around the restatement announcement is significantly positive in both periods. *VOLUME* is significantly lower after SOX, even though investor disagreement would tend to generate more volume. Part of the reason for the lower post-SOX *VOLUME* is that post-SOX restatements are less newsworthy, as evidenced by the less negative announcement returns. Later I use regression (2) to determine whether *VOLUME* is higher after SOX, holding newsworthiness constant.

Consistent with the less negative returns after SOX, post-SOX restatements are less egregious across many dimensions. Post-SOX restatements' mean impact on earnings as a percentage of assets (*MAG*) is only about half that of pre-SOX restatements (−2.3 percent post-SOX versus −4.1 percent pre-SOX). Post-SOX restatements are less likely to involve *FRAUD* (21.4 percent post-SOX versus 34.9 percent pre-SOX), more likely to be prompted by management (*MGT-_PROMPTS*, 71.4 percent post-SOX versus 47.2 percent pre-SOX), and less likely to involve *CORE* earnings items (23.3 percent post-SOX versus 39.8 percent pre-SOX).[16]

Announcement Return Results

Table 3 presents coefficient estimates from the announcement return regression (1) using OLS with heteroscedasticity-consistent standard errors (White 1980). One specification uses short window returns as the dependent variable and the other uses episode returns. The episode return regression has fewer observations because it excludes observations that are missing I/B/E/S data for any earnings announcement in the episode window. In both regressions, the coefficient on *POSTSOX* is significantly positive, statistically and economically, suggesting that investors react less negatively to restatements after SOX even after controlling for the decline in egregiousness. In the short window (episode window) regression, the *POSTSOX* coefficient is 0.0464 (0.0597), suggesting that mean returns are 4.64 (5.97) percentage points higher after SOX, holding other factors constant.[17]

The control variables perform similarly across both regressions. As expected, the restatement earnings impact (*MAG*) is significantly related to returns, and the strength of this relation diminishes with large earnings impacts (i.e., the coefficient on $MAG * |MAG|$ is significantly negative). Returns are significantly more negative when the restatement lacks an initial magnitude disclosure (*NO_INITIAL_MAG*), involves long delays in releasing the magnitude (*LONG_MAG_DELAY*), and involves *FRAUD*, *CORE* items, or *MULTIPLE* errors. Unexpectedly, *QUARTERLY_ONLY* restatements are significantly associated with more negative announcement returns. Perhaps the market suspects that these restatements result from managers intentionally shifting earnings across

---

[15] Huang (2010) reports that mean announcement returns for her restatement sample begin to decline in 2001. In contrast, in my sample, mean announcement returns (as opposed to median episode returns) do not exhibit a noticeable decline until 2003 (untabulated). The differing results may be due to differences in samples; Huang (2010) uses the Palmrose and Scholz (2004b) sample from 1995 to 2000 and the Audit Analytics sample from 2000 to 2006. The focus of my study is not to pinpoint when market reactions begin to decline; rather, I focus on the period when investors had many restatements to track, which begins in 2002.

[16] The *FRAUD* rate for pre-SOX restatements in my sample is lower than the near 50 percent rate that Hennes et al. (2008) report for their pre-SOX sample, but their sample consists of 1997 and 1998 only. In these years, my sample has a *FRAUD* rate of 47 percent, similar to Hennes et al.'s (2008).

[17] *POSTSOX* remains statistically and economically significant after excluding firms with low stock prices (less than $5 per share, two trading days prior to the restatement announcement), and excluding influential observations (observations for which the dfbetas statistic for *POSTSOX* has an absolute value higher than the $2 / (n^{1/2})$ cutoff recommended by Belsley et al. 1980). Because *FRAUD* is a crucial determinant of announcement returns (Hennes et al. 2008), I partition the sample into fraud and non-fraud observations and rerun the regressions. *POSTSOX* remains statistically and economically significant in both partitions.

TABLE 3

Tests of Announcement Returns

	Predicted Sign	Announcement Returns (*RETURN_ANNOUNCE*)	Episode Returns (*RETURN_EPISODE*)
Intercept	?	−0.0075	0.0157
		(0.0315)	(0.0494)
POSTSOX	?	0.0464 **	0.0597**
		(0.0098)	(0.0143)
Restatement Controls			
MAG	+	0.4160 *	0.9180**
		(0.1791)	(0.3104)
POSITIVE_MAG	+	0.0047	−0.0169
		(0.0076)	(0.0106)
MAG × *POSITIVE_MAG*	−	−0.1589	−0.7204
		(0.2427)	(0.4566)
MAG × |*MAG*|	−	−0.6567 *	−1.2072*
		(0.3857)	(0.6917)
NO_INITIAL_MAG	−	−0.0273 **	−0.0425**
		(0.0089)	(0.0109)
LONG_MAG_DELAY	−	−0.0197 *	−0.0364*
		(0.0097)	(0.0160)
FRAUD	−	−0.0518 **	−0.0756**
		(0.0105)	(0.0188)
FRAUD_LOWLEVEL	−	−0.0122	0.0024
		(0.0175)	(0.0275)
MGT_PROMPTS	+	−0.0060	−0.0002
		(0.0075)	(0.0113)
CORE	−	−0.0203 *	−0.0218*
		(0.0088)	(0.0126)
MULTIPLE	−	−0.0203 *	−0.0275*
		(0.0096)	(0.0162)
QUARTERLY_ONLY	+	−0.0407 **	−0.0581**
		(0.0092)	(0.0125)
General Controls			
RETURN_PRIOR	−	−0.0144	−0.0150
		(0.0120)	(0.0191)
FIRMSIZE	+	−0.0005	−0.0003
		(0.0019)	(0.0030)
VOLATILITY_INDEX	−	−0.0005	−0.0006
		(0.0006)	(0.0008)
ESURPRISE	+	0.2305	
		(0.3384)	
ESURPRISE_EPISODE	+		0.4022**
			(0.1610)
MBE	+	0.0335 **	
		(0.0087)	

(continued on next page)

TABLE 3 (continued)

	Predicted Sign	Announcement Returns (*RETURN_ANNOUNCE*)	Episode Returns (*RETURN_EPISODE*)
Adj. R^2		20.3%	22.4%
n Pre-SOX		407	352
n Post-SOX		819	716
n		1,226	1,068

*, ** Denote significance at the 5 percent and 1 percent levels, respectively (one-tailed if sign is in the predicted direction, two-tailed otherwise).

Models are estimated using OLS with heteroscedasticity-consistent standard errors (White 1980). Standard errors are presented in parentheses below coefficient estimates. The dependent variable is size-adjusted stock returns measured over either a "short" or "episode" window around the restatement announcement (*RETURN_ANNOUNCE* or *RETURN_EPISODE*).

See the Appendix for variable descriptions.

quarters to smooth trends or beat forecasts. The earnings surprise proxies are the only significant general control variables. The meet-or-beat indicator (*MBE*) is significantly related to short window returns, and the continuous measure of earnings surprises during the restatement episode (*ESURPRISE_EPISODE*) is significantly related to episode returns.[18]

Robustness of the *POSTSOX* Effect

In untabulated analysis, I test for nonlinearities involving *MAG* that might explain the observed *POSTSOX* effect by interacting *MAG* with the restatement characteristics that were statistically significant. The only statistically significant interaction is *MAG* * *FRAUD*; its positive coefficient suggests that the relation between returns and earnings impact is even stronger for fraudulent restatements. The coefficient on *POSTSOX* remains significantly positive, statistically and economically.

The post-SOX period contains four events that increased scrutiny on accounting reports, possibly leading to more technical restatements that spur less negative reactions. I include indicator variables to test whether the *POSTSOX* effect can be explained by restatements announced during these events: (1) re-audits of former Arthur Andersen clients after Andersen's collapse, (2) the first time managers had to certify their financial statements under SEC Order No. 4-460, (SEC 2002), (3) the first time managers had to certify their financial statements under SOX Section 302, and (4) the first time managers had to assess internal controls under SOX Section 404. The only statistically significant event indicator is the one related to SOX 404, and it is significant only in the short window regression. The *POSTSOX* coefficient remains statistically and economically significant in both regressions.

At this point, the analysis suggests that either investors under-react to post-SOX restatements, or that investor reaction is efficient but the model fails to control for changes in value-relevant factors after SOX. The next step is to determine whether prices drift negatively over the long run, which would indicate that the less negative initial reactions are under-reactions.

[18] None of the regression inferences appear to be affected by collinearity. *MAG* and *MAG* * |*MAG*| have variance inflation factors (VIFs) around 12 and 10, which is not surprising given that the regression has several terms involving *MAG*. Der and Everitt (2002) consider VIFs higher than 10 cause for concern. Excluding *MAG* * |*MAG*| does not change inferences. All other variables have VIFs around 2 or below.

To focus the price drift tests on the subsets of restatements most likely to be mispriced, I use the episode return regression to determine whether particular types of restatements are responsible for the less negative post-SOX reactions. I interact *POSTSOX* with each independent variable one at a time, rerun the regression, and note which interactions reduce the coefficient magnitude on the *POSTSOX* intercept by two or more standard deviations. Based on this criterion, restatements involving *FRAUD*, long episodes (*LONG_MAG_DELAY*), and no initial earnings impact disclosure (*NO_INITIAL_MAG*) appear partially responsible for the less negative post-SOX reactions.

Long-Run Price Drift Results

I test for drift over six-month, one-year, and two-year horizons beginning at two different points: two days after the initial restatement announcement and two days after the earnings impact announcement. I use the earnings impact announcement as an alternative starting point because the initial restatement announcement often does not provide investors with full information about the restatement. I measure drift using restatement firms' median abnormal buy-and-hold return over the horizon, and consider the return statistically significant if it falls in the top or bottom 2.5 percentile of the empirical distribution of pseudo-portfolio median abnormal buy-and-hold returns. This procedure is essentially a two-tailed test conducted at the 5 percent level of significance.

Table 4, Panel A shows the distribution of the pseudo-portfolio median abnormal returns by SOX period at the three horizons. The means of the six-month, one-year, and two-year pre-SOX (post-SOX) distributions are −4.8, −11.4, and −17.1 percent (−3.5, −7.8, and −13.4 percent), respectively, indicating that negative bias exists in both SOX periods and is larger before SOX. The pseudo-portfolio approach controls for this negative bias because the ranking of the restatement portfolio return among the pseudo-portfolio returns determines statistical significance. In other words, if the restatement portfolio exhibits no drift other than that caused by the negative bias that affects all portfolios, then the restatement portfolio's drift would not fall in the extreme lower tail of the pseudo-portfolio distribution.[19]

Table 4, Panel B shows the median abnormal returns (i.e., drifts) of the restatement firms. Negative drifts in the post-SOX period would indicate that the less negative post-SOX announcement reactions documented in the previous section represent under-reactions to restatements. Panel B shows some evidence of negative drift following post-SOX restatements, but only when the starting point is immediately after the initial announcement (see second column of drifts); the full post-SOX sample has a significantly negative drift over the six months after the initial announcement (−5.3 percent), but not over one or two years. Turning to the post-SOX subsamples, restatements with long episodes (*LONG_MAG_DELAY*) have significantly negative drifts over the six months after the initial announcement (−7.5 percent), and fraudulent restatements have significantly negative drifts over the six months and one year after the initial announcement (−10.4 and −14.4 percent, respectively).[20] In contrast, the drifts beginning after the earnings impact announcement are smaller in magnitude and not statistically significant for the full post-SOX sample or subsamples (see fourth column). These findings suggest that the drifts beginning after the initial announcement are caused by firms slowly releasing restatement information rather than by investors mispricing released information.

---

[19] Whereas the median pseudo-portfolio returns are negatively biased, the mean pseudo-portfolio returns are unbiased, i.e., when computing the *mean* abnormal return for each of the 1,000 pseudo-portfolios, the distribution of the 1,000 means has a mean near zero (untabulated). This indicates that the negative bias in the median pseudo-portfolio returns occurs because the median is not as affected by large positive returns among a relatively small number of firms. I focus on medians to limit the influence of these firms.

[20] I assess the significance of the drift in each subsample using pseudo-portfolios consisting of only the control firms matched to the restatement firms in each particular subsample.

TABLE 4
Tests of Long-Run Price Drifts

Panel A: Empirical Distributions of the 1,000 Pseudo-Portfolio Median Abnormal Buy-and-Hold Returns (cumulation begins after the restatement's earnings impact is announced)

	Mean	Std. Error	Min	Percentiles 10th	25th	50th	75th	90th	Max
Pre-SOX									
6 months	−4.8%	1.8%	−10.3%	−7.1%	−5.9%	−4.8%	−3.5%	−2.5%	1.4%
1 year	−11.4%	2.9%	−19.5%	−15.1%	−13.5%	−11.5%	−9.4%	−7.5%	−0.9%
2 years	−17.1%	3.9%	−31.1%	−21.9%	−19.5%	−17.1%	−14.4%	−12.3%	−3.8%
Post-SOX									
6 months	−3.5%	0.9%	−6.2%	−4.7%	−4.0%	−3.5%	−2.9%	−2.4%	−1.2%
1 year	−7.8%	1.4%	−12.2%	−9.7%	−8.7%	−7.8%	−6.9%	−6.0%	−4.1%
2 years	−13.4%	2.1%	−20.8%	−16.3%	−14.8%	−13.3%	−11.9%	−10.7%	−7.0%

Panel B: Median Abnormal Buy-and-Hold Returns for Restatement Firms

	Return Cumulation Begins Two Days After: Initial Restatement Announcement Pre-Sox	Initial Restatement Announcement Post-SOX	Earnings Impact Announcement Pre-SOX	Earnings Impact Announcement Post-SOX
Full Sample				
Horizon				
6 months	**−16.1%****	**−5.3%***	**−11.6%****	**−4.4%**
1 year	**−27.9%****	**−8.4%**	**−22.1%****	**−6.7%**
2 years	**−39.3%****	**−11.9%**	**−38.7%****	**−11.1%**
n	422	792	416	783

(continued on next page)

Panel B: Median Abnormal Buy-and-Hold Returns for Restatement Firms

	Return Cumulation Begins Two Days After:			
	Initial Restatement Announcement		Earnings Impact Announcement	
	Pre-Sox	Post-SOX	Pre-SOX	Post-SOX
Sample Partitions				
No Initial Magnitude (*NO_INITIAL_MAG* = 1)				
6 months	**−20.2%****	**−4.9%**	**−11.2%****	**−3.8%**
1 year	**−30.6%****	**−8.1%**	**−20.9%****	**−5.8%**
2 years	**−40.4%****	**−9.9%**	**−39.4%****	**−10.1%**
n	212	478	206	469
Long Magnitude Delay (*LONG_MAG_DELAY* = 1)				
6 months	**−21.8%****	**−7.5%****	−9.3%	−4.9%
1 year	**−30.1%****	**−12.0%**	−9.6%	−3.9%
2 years	**−40.6%****	**−16.3%**	**−33.1%****	**−13.2%**
n	131	293	125	284
Fraudulent (*FRAUD* = 1)				
6 months	**−28.4%****	**−10.4%****	**−23.9%****	**−5.7%**
1 year	**−30.8%****	**−14.4%****	**−25.1%****	**−6.7%**
2 years	**−47.2%****	**−14.7%**	**−45.5%****	**−15.6%**
n	130	161	124	153
Restatement Egregiousness Factor—Least Egregious Quartile				
6 months	−9.4%	−5.7%	−9.4%	−5.5%
1 year	−5.9%	−7.8%	−5.9%	−7.8%
2 years	−27.5%	−12.6%	−25.6%	−12.6%
n	98	194	98	194
Restatement Egregiousness Factor—Most Egregious Quartile				
6 months	**−27.3%****	**−11.7%****	**−17.7%****	**−6.2%**
1 year	**−35.3%****	**−16.5%****	−22.0%*	−7.4%
2 years	**−44.8%****	**−20.2%**	**−43.8%****	**−16.0%**
n	118	181	112	173

(continued on next page)

Panel B: Median Abnormal Buy-and-Hold Returns for Restatement Firms

	Return Cumulation Begins Two Days After:			
	Initial Restatement Announcement		Earnings Impact Announcement	
	Pre-Sox	Post-SOX	Pre-SOX	Post-SOX
Announcement Returns—Most Positive Quartile				
6 months	−14.6%	−2.9%	−13.4%	−2.3%
1 year	**−33.3%****	**−5.4%**	−26.4%	−3.9%
2 years	−32.4%	−8.1%	−32.4%	−5.4%
n	85	212	84	212
Announcement Returns—Most Negative Quartile				
6 months	**−25.4%****	**−7.9%**	**−15.7%****	**−5.1%**
1 year	**−32.3%****	**−11.2%**	**−28.6%****	**−8.7%**
2 years	**−47.9%****	**−22.8%**	**−46.7%****	**−21.1%**
n	168	147	163	141
Unexplained Announcement Return—Highest Quartile (less negative initial reactions than expected)				
6 months	−15.4%**	−7.9%**	−12.7%*	−4.0%
1 year	**−35.7%****	**−12.3%***	**−23.9%***	**−7.6%**
2 years	**−35.9%****	**−9.0%**	**−32.4%***	**−6.1%**
n	87	169	86	165
Unexplained Announcement Return—Lowest Quartile (more negative initial reactions than expected)				
6 months	**−30.2%****	**−6.8%**	**−22.1%****	**−5.7%**
1 year	**−38.5%****	**−9.9%**	**−32.5%****	**−10.8%**
2 years	**−60.5%****	**−17.6%**	**−50.0%****	**−19.6%**
n	93	179	91	177

(continued on next page)

Panel C: Fama-French Abnormal Monthly Median Returns for Restatement Firms

	Return Cumulation Begins Two Days After:			
	Initial Restatement Announcement		**Earnings Impact Announcement**	
	Pre-Sox	**Post-SOX**	**Pre-SOX**	**Post-SOX**
Full Sample				
Horizon				
6 months	**−3.2%****	**−0.7%***	**−2.4%****	**−0.8%****
1 year	**−2.4%****	**−0.6%***	**−2.1%****	**−0.5%****
2 years	**−2.4%****	**−0.5%***	**−2.2%****	**−0.5%****
Sample Partitions				
No Initial Magnitude (*NO_INITIAL_MAG* = 1)				
6 months	**−4.9%****	**−0.6%**	**−2.4%****	**−0.8%***
1 year	**−3.0%****	**−0.5%***	**−2.2%****	**−0.5%***
2 years	**−2.7%****	**−0.5%****	**−2.3%****	**−0.5%****
Long magnitude delay (*LONG_MAG_DELAY* = 1)				
6 months	**−5.0%****	**−1.0%***	−2.0%**	−1.2%*
1 year	−2.3%**	−0.8%*	−0.9%	−0.8%**
2 years	**−2.4%****	**−0.7%****	**−2.0%****	**−0.6%****
Fraudulent (*FRAUD* = 1)				
6 months	**−5.6%****	**−1.4%***	−2.6%*	−2.0%**
1 year	−2.9%*	−1.2%**	−2.1%*	−1.5%**
2 years	**−2.7%****	**−1.2%****	**−2.2%****	**−1.0%****
Restatement Egregiousness Factor—Least Egregious Quartile				
6 months	−0.8%	−1.2%**	−0.8%	−1.2%**
1 year	−1.2%	−0.8%**	−1.2%	−0.8%**
2 years	−1.4%**	−0.6%**	−1.4%**	−0.6%**
Restatement Egregiousness Factor—Most Egregious Quartile				
6 months	**−6.4%****	**−1.3%***	−2.6%**	−1.6%**
1 year	**−3.8%****	**−1.3%****	−1.9%*	−1.2%**
2 years	**−2.7%****	**−1.0%****	**−2.2%****	**−0.9%****
Announcement Returns—Most Positive Quartile				
6 months	−2.2%	−0.7%	−1.0%	−0.9%
1 year	−2.1%**	−0.5%	−1.8%*	−0.6%*
2 years	**−2.4%****	**−0.5%***	**−2.2%****	**−0.5%***

(continued on next page)

Panel C: Fama-French Abnormal Monthly Median Returns for Restatement Firms

	Return Cumulation Begins Two Days After:			
	Initial Restatement Announcement		**Earnings Impact Announcement**	
	Pre-Sox	**Post-SOX**	**Pre-SOX**	**Post-SOX**
Announcement Returns—Most Negative Quartile				
6 months	**−5.3%****	**−1.0%**	**−4.0%****	**−1.6%***
1 year	**−4.2%****	**−1.0%**	**−3.7%****	**−1.3%***
2 years	**−3.6%****	**−1.1%****	**−3.5%****	**−1.1%****
Unexplained Announcement Return—Highest Quartile (less negative initial reactions than expected)				
6 months	**−5.9%****	**−0.8%**	−2.9%*	−0.6%
1 year	**−3.1%****	**−0.5%**	−1.9%*	−0.2%
2 years	**−2.1%****	**−0.4%**	**−1.9%****	**−0.3%**
Unexplained Announcement Return—Lowest Quartile (more negative initial reactions than expected)				
6 months	**−5.0%****	**−0.8%**	**−4.5%****	**−1.7%****
1 year	**−4.0%****	**−0.9%***	**−3.5%****	**−1.2%****
2 years	**−3.4%****	**−1.1%****	**−3.3%****	**−1.3%****

Abnormal returns for each firm are computed by taking the firm's buy-and-hold return minus the buy-and-hold return of an equally weighted benchmark portfolio matched on size and book-to-market.

Panel A:

To form a single pseudo-portfolio, I match each restatement firm randomly with replacement to a non-restatement firm in the same size/book-to-market benchmark portfolio. Then I compute an abnormal return for each firm in the pseudo-portfolio as described above, and find the median abnormal return of the pseudo-portfolio. This process is repeated 1,000 times to form the empirical distributions of median returns shown above.

Panel B:

*, ** Denote that the value lies in the top or bottom 2.5 and 0.5 percentiles of the empirical distribution, respectively.

To identify a statistically significant difference from 0, I rank the median buy-and-hold abnormal return of the restatement firms within the empirical distribution of pseudo-portfolio returns. This corresponds to a two-tailed test at commonly accepted levels of significance. In each subsample, significance is assessed using pseudo-portfolios consisting only of the control firms matched to the restatement firms in that subsample.

To identify a statistically significant difference across SOX periods, I use an empirical distribution formed by computing the difference between the pre- and post-SOX median return for each pseudo-portfolio, and then rank the difference in median return across pre- and post-SOX restatement firms within the empirical distribution. Bold denotes that the difference between pre- and post-SOX returns lies in the top or bottom 2.5 percentile of the empirical distribution. This corresponds to a two-tailed test commonly accepted levels of significance.

Panel C:

(continued on next page)

*, ** Denote that the intercept statistically differs from 0 at the 5 percent and 1 percent levels, two-tailed, respectively. Bold denotes that the pre- and post-SOX intercepts statistically differ from each other at the 5 percent level, two-tailed.

Panel C reports the estimated intercepts from Fama and French (1993) calendar-time regressions of median monthly restatement firm returns on the three Fama-French risk factors. The regression is estimated using weighted least squares, with the weight corresponding to the number of firms in the restatement portfolio. Significance is assessed using heteroscedasticity-consistent standard errors (White 1980). To statistically assess the difference between pre- and post-SOX intercepts, I take the difference between the estimated intercepts and divide by the square root of the sum of the two intercepts' estimated variances, which is asymptotically standard normal assuming independent pre- and post-SOX samples.

I also rank restatements by egregiousness, announcement return, and unexplained announcement return, testing for drifts in the top and bottom quartiles of each variable. Restatement egregiousness is the score from a principle components analysis of the five restatement characteristics that were significant in the predicted direction in the announcement returns regression: *MAG*, *NO_INITIAL_MAG*, *LONG_MAG_DELAY*, *FRAUD*, and an indicator variable capturing whether the restatement involves *CORE* or *MULTIPLE* items.[21] I compute quartiles for the egregiousness factor and announcement return on the sample as a whole; they are not computed separately for pre- and post-SOX periods. For example, in order for a post-SOX restatement to be in the most egregious quartile, it must be among the most egregious of pre- and post-SOX restatements combined. I compute unexplained announcement return by running the short window announcement return regression (1) on the pre-SOX sample only, using the estimated coefficients to predict announcement returns for both the pre- and post-SOX samples, and computing the difference between actual and predicted.

For post-SOX restatements, the most egregious quartile and the quartile with the most positive unexplained announcement returns have significantly negative six-month and one-year drifts following the initial announcement. None of the quartiles have significant drifts following the earnings impact announcement. To summarize, the post-SOX period contains some evidence of negative drift following initial announcements, but no evidence of drifts once investors have relatively complete information about the restatements.

In contrast, the pre-SOX period contains strong evidence of negative drifts even after investors have relatively complete information about the restatements. For the full pre-SOX sample, drifts at six months, one year, and two years following the earnings impact announcement are −11.6, −22.1, and −38.7 percent, respectively, all significant at the 1 percent level (see third column). The negative drifts are consistent with investor under-reaction documented for many other types of public news events, including large quarterly earnings surprises, stock splits, stock repurchases, and dividend initiations and omissions (Daniel et al. 1998).

Comparing across periods, the negative drifts in the pre-SOX period are larger than those in the post-SOX period. The pairs of drifts that statistically differ across SOX periods are shown in bold in Panel B.[22] In the full sample, pre-SOX drifts are significantly more negative at all horizons. In the sample partitions, pre-SOX drifts are almost always more negative, and the difference is usually statistically significant. The pre-SOX drifts tend to be significantly more negative even in the quartiles of restatements that are the most egregious and that have the most negative initial reactions. This shows that the more negative pre-SOX drifts are not simply due to the more egregious nature of pre-SOX restatements; rather, investors appear to price even the most egregious restatements more efficiently after SOX.

Robustness Checks for Buy-and-Hold Drifts

Pre-SOX restatements may have more negative drifts because pre-SOX restatement episodes are shorter; i.e., in the pre-SOX period, firms are more likely to announce the restatement's

[21] The components load on a single factor using the Kaiser criterion—i.e., only the first factor has an eigenvalue greater than 1. The factor explains 45 percent of the total variation in the variables. The standardized scoring coefficients for *MAG*, *FRAUD*, *NO_INITIAL_MAG*, *LONG_MAG_DELAY*, and the *CORE* or *MULTIPLE* indicator are −0.194, 0.324, 0.333, 0.363, 0.240, respectively.

[22] To determine whether the difference in pre- and post-SOX drifts is statistically significant, I take each of the 1,000 pseudo-portfolios and subtract the pre-SOX median drift from the post-SOX median drift. This creates a distribution of 1,000 differences in pre- and post-SOX median drifts. Then I determine whether the difference in pre- and post-SOX median drift among restatement firms falls in the top or bottom 2.5 percentile of the empirical distribution. Note that this test assesses whether pre-SOX restatements have more negative drifts than post-SOX restatements after controlling for the larger negative bias that exists in pre-SOX drifts (Table 4, Panel A).

earnings impact at the time of the initial announcement or shortly thereafter, giving investors less time to process and react to the information. Pre-SOX episodes are shorter than post-SOX episodes by a mean (median) of 13.5 (8.5) days. To investigate the role of shorter episodes, I compute a new set of drifts for pre-SOX restatements that do not begin until 17 days after the earnings impact announcement. Then I compare this set to the original set of post-SOX drifts that begin only two days after the earnings impact announcement. In the full sample, the pre-SOX drifts remain significantly negative and significantly more negative than the post-SOX drifts, and the same tends to be true of the sample partitions (untabulated).

Next, I repeat the analysis of Table 4 using mean drifts instead of median drifts, and focus on the drifts beginning after the earnings impact announcement (untabulated). In both the pre- and post-SOX periods, the magnitudes of the drifts become smaller, and rarely significantly differ from zero or differ across SOX periods. In the full sample, the pre-SOX (post-SOX) drifts at six-month, one-year, and two-year horizons are −6.0, −7.8, and −2.1 percent (−0.6, 0.2, and 2.4 percent), respectively. None of these drifts are significantly different across SOX periods, and only the pre-SOX, six-month drift significantly differs from zero. The decrease in drift magnitude when using means indicates that a small number of firms have large positive drifts, which offset the negative drifts experienced by most of the restatement firms. In other words, the lack of significant mean drift hides the fact that, before SOX, investors typically under-reacted to restatements.[23]

I also test the influence of small firms, which are more likely to be mispriced because of illiquidity or low institutional following. I rerun the median and mean drift analyses using just the firms in the largest 9 of the 14 size partitions. The magnitudes and significance levels of the drifts are similar to those in the original analyses, and inferences are unchanged.

Fama-French Tests of Price Drifts

An alternative approach for detecting long-run price drifts is the Fama and French (1993) calendar-time regression. This approach tests whether monthly returns are persistently abnormal over the horizon, rather than whether the overall return for the horizon is abnormal (Lyon et al. 1999). The dependent variable for the Fama-French regression is a single time-series of monthly restatement portfolio returns. A firm enters the portfolio in the first calendar month following its restatement announcement and remains in the portfolio for 6, 12, or 24 months, depending on the testing horizon. To compute the portfolio return in each month t, I take the cross-sectional median of the returns of the firms in the portfolio (R_p), and subtract the one-month Treasury rate (R_f). I regress this monthly portfolio return on the three Fama-French risk factors, allowing the intercept (α) to capture the abnormal portion of the monthly return. The three risk factors are the value-weighted market return (R_m) less one-month Treasury rate, the return on a hedge portfolio of small and large stocks (SMB), and the return on a hedge portfolio of high and low book-to-market stocks (HML):[24]

$$R_{pt} - R_{ft} = \alpha + \beta_1(R_{mt} - R_{ft}) + \beta_2 SMB_t + \beta_3 HML_t + \varepsilon_t. \tag{5}$$

I estimate the model using weighted least-squares, with the weight corresponding to the number of firms in the restatement portfolio in month t, and use heteroscedasticity-consistent standard errors.

[23] Not surprisingly, I find that both median and mean *unsigned* price drifts following the earnings impact date are higher before SOX, significantly so at all horizons for medians and at the 24-month horizon for means. I assess significance by computing the difference between pre- and post-SOX unsigned drift for each pseudo-portfolio, forming a distribution of 1,000 differences. Then I determine where the difference between pre- and post-SOX unsigned drift for the restatement portfolio lies in this distribution.

[24] I thank Kenneth French for providing the factors on his website: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/f-f_factors.html.

Table 4, Panel C shows estimates of monthly α over six-month, one-year, and two-year horizons. Focusing on the period beginning after the earnings impact date (columns 3 and 4), estimates of α tend to be significantly negative for both pre- and post-SOX restatements in the full sample and the partitions. This finding contrasts with the results for buy-and-hold returns, which were not significantly negative for post-SOX restatements. However, consistent with the buy-and-hold return results, pre-SOX α estimates are often significantly more negative than post-SOX α estimates.[25] In the full sample, pre-SOX α estimates over six-month, one-year, and two-year horizons are -2.4, -2.1, and -2.2 percent, and the corresponding post-SOX α estimates are -0.8, -0.5, and -0.5 percent, respectively. Each of these α estimates is significantly negative, and the pre-SOX α estimates are statistically more negative than the post-SOX α estimates at all horizons. Converting these monthly α estimates to accumulated six-month, one-year, and two-year median returns, the results are -14.4, -25.2, and -52.8 percent for pre-SOX restatements and -4.8, -6.0, and -12.0, respectively, for post-SOX restatements.[26] The α estimates for the sample partitions exhibit similar patterns. In summary, both the buy-and-hold and Fama-French results for median returns suggest that investors typically price restatements more efficiently after SOX, despite the larger number of restatements to process.

For robustness, in untabulated tests I estimate α using mean instead of median monthly restatement firm returns. Similar to the results for mean buy-and-hold returns from the previous section, estimates of α tend to be statistically insignificant from zero and never significantly differ across SOX periods. I also rerun the Fama-French analyses after restricting the sample to firms in the largest 9 of the 14 size partitions, and find results similar to those of the full sample.

Volume Results

Even if investor confusion does not cause mispricing, confusion could lead to disagreements among investors that spur more trading volume relative to contemporaneous price changes. Table 5 shows the results of regressing abnormal trading *VOLUME* in the three days around the restatement announcement on *POSTSOX* and control variables. The *POSTSOX* coefficient is positive, as predicted, but not statistically different from zero, providing no reliable evidence of increased investor disagreement after SOX. The coefficient on absolute abnormal return (*AB_RETURN*) is significantly positive, confirming that volume increases in the news of the announcement. The interaction between *AB_RETURN* and *POSTSOX* is insignificant, suggesting that the relation between volume and absolute return is similar across the two SOX periods. Because *POSTSOX* appears in the regression as a main term and as an interaction with *AB_RETURN*, I also evaluate the combined marginal effect of the *POSTSOX* terms on volume when *AB_RETURN* is at its mean. The combined effect of the *POSTSOX* terms is negative and insignificant, contrary to predictions of higher volume in the post-SOX period.[27]

Turning to the remaining control variables, the *AB_RETURN* * *POSITIVE_RETURN* interaction is significantly negative, suggesting that volume is not as responsive to restatement news

[25] To statistically assess the difference between pre- and post-SOX α estimates, I take the difference between the α estimates and divide by the square root of the sum of the two estimated α variances, which is asymptotically standard normal, assuming independent pre- and post-SOX samples.

[26] To convert to an accumulated return, I multiply the six-month α by 6, the one-year α by 12, and the two-year α by 24.

[27] The combined effect is -0.002, computed as: 0.0032 *POSTSOX* coefficient + (-0.0553 interaction coefficient x 0.094 mean *AB_RETURN*). The *POSTSOX* main effect and combined effect remain statistically insignificant in two alternative specifications. The first specification limits the sample to firms that have stock prices of at least $5 per share. The second specification does not adjust the dependent variable for the firm's normal level of volume, i.e., the specification uses only $TO_{restate,i}$ from Equation (4) as the dependent variable. I run this second specification because if investors are more confused about poorly performing firms, then adjusting for the firm's normal level of volume could hide the effect of interest. In all specifications, multicollinearity does not appear to affect inferences, as variance inflation factors are low.

TABLE 5

Tests of Volume around Restatement Announcements

	Predicted Sign	Dependent Variable *VOLUME*
Intercept	?	−0.0397**
		(0.0090)
POSTSOX	+	0.0032
		(0.0036)
Restatement Controls		
AB_RETURN	+	0.1929**
		(0.0299)
AB_RETURN × *POSTSOX*	?	−0.0553
		(0.0319)
POSITIVE_RETURN	?	0.0068**
		(0.0019)
AB_RETURN × *POSITIVE_RETURN*	?	−0.1298**
		(0.0172)
AB_MAG	+	0.0118
		(0.0157)
POSITIVE_MAG	−	0.0019
		(0.0020)
NO_INITIAL_MAG	+	0.0015
		(0.0019)
LONG_MAG_DELAY	+	−0.0029
		(0.0027)
FRAUD	+	0.0065**
		(0.0028)
FRAUD_LOWLEVEL	+	−0.0083**
		(0.0027)
MGT_PROMPTS	−	−0.0012
		(0.0019)
CORE	+	0.0034*
		(0.0021)
MULTIPLE	+	0.0025
		(0.0021)
QUARTERLY_ONLY	−	−0.0001
		(0.0025)
General Controls		
AB_RETURN_PRIOR	−	0.0022
		(0.0067)
FIRMSIZE	?	0.0026**
		(0.0005)
VOLATILITY_INDEX	+	−0.0002*
		(0.0001)
AB_ESURPRISE	+	0.0684
		(0.0629)
MBE	?	0.0031
		(0.0021)

(continued on next page)

TABLE 5 (continued)

	Predicted Sign	Dependent Variable *VOLUME*
Adj. R^2		33.8%
n Pre-SOX		407
n Post-SOX		819
n		1,226

*, ** Denote significance at the 5 percent and 1 percent levels, respectively (one-tailed if sign is in the predicted direction, two-tailed otherwise).

Models are estimated using OLS with heteroscedasticity-consistent standard errors (White 1980). Standard errors are presented in parentheses below coefficient estimates. The dependent variable is abnormal *VOLUME* over days (−1, +1) relative to the restatement announcement.

See the Appendix for variable descriptions.

when the news is good. Restatements involving *FRAUD* and *CORE* items are significantly associated with higher volume, although restatements that involve fraud at lower levels of the firm (*FRAUD_LOWLEVEL*) are significantly associated with lower volume. Volume is also significantly increasing in *FIRMSIZE* and decreasing in the implied level of market volatility (*VOLATILITY_INDEX*).

VI. CONCLUSION

Pricing and volume tests provide little evidence that investors are confused by post-SOX restatements. Initial reactions to post-SOX restatements are less negative than those of pre-SOX restatements even after attempting to control for the less egregious nature of post-SOX restatements. While this finding could be consistent with investor confusion, there is little evidence of negative long-run price drifts unique to the post-SOX period, which would occur if the less negative initial reactions represented under-reactions by confused investors. Far from appearing confused, that market appears to have become significantly more efficient at pricing restatements after SOX, as the price drifts are often smaller after SOX than before.

Prices reflect the aggregate judgment of investors. Although aggregate judgment appears to improve after SOX, it is possible that confusion among individual market participants could lead to more disagreements over valuations. I test for increased disagreement by examining trading volume incremental to price movements around restatement announcements, but do not find evidence of an increase in disagreement after SOX.

These results alleviate concerns that the dramatic increase in restatements after SOX has confused investors. However, there may be other reasons to curtail restatements, which future research can address. While this study focuses on the capital market consequences for the restating firms themselves, the flood of restatements after SOX could affect public confidence in the capital markets as a whole. Future studies could examine macro-level effects of restatements, or other concerns of regulators such as the cost to preparers.

From a standard-setting perspective, the results suggest that investors were able to adapt as more firms revised financial statements to correct errors. To better understand investors' capacity to process other types of financial statement revisions, future research could examine revisions due to changes in accounting principle or changes in reporting entity. In 2005, firms began revising

prior financial statements for changes in accounting principle, so evidence on how investors adapted would help standard-setters decide whether to expand retrospective treatment to more types of accounting changes.

APPENDIX
VARIABLE DESCRIPTIONS

Panel A: Dependent Variables

RETURN_ANNOUNCE	Size-adjusted buy-and-hold stock return over days (−1, +1) relative to the initial restatement announcement, including any CRSP delisting return. Winsorized at 1 and 99 percent.
RETURN_EPISODE	Size-adjusted buy-and-hold stock return from one day before the initial restatement announcement to one day after the announcement of the restatement's earnings impact, including any CRSP delisting return. Winsorized at 1 and 99 percent.
VOLUME	Abnormal volume over days (−1, +1) relative to the initial restatement announcement, computed similarly to Garfinkel and Sokobin (2006).

Panel B: Explanatory Variable of Interest

POSTSOX	Equals 1 if the restatement is announced in the month SOX was passed (July 2002) or after, 0 otherwise.

Panel C: Other Explanatory Variables

AB_ESURPRISE	Absolute value of *ESURPRISE*.
AB_MAG	Absolute value of *MAG*.
AB_RETURN_PRIOR	Absolute value of *RETURN_PRIOR*.
AB_RETURN	Absolute value of *RETURN_ANNOUNCE*.
CORE	Equals 1 if the restatement affects components of pre-tax operating income other than non-cash items like depreciation, amortization, and equity-based compensation, 0 otherwise.
ESURPRISE	If earnings are announced in the three-day window around the restatement announcement, then *ESURPRISE* is the difference between actual earnings per share and the most recent consensus analyst forecast from I/B/E/S, scaled by stock price one day before the earnings announcement. Winsorized at 1 and 99 percent. If earnings are not announced in the three-day window around the restatement announcement, then *ESURPRISE* equals 0.
ESURPRISE_EPISODE	The sum of all the individual *ESURPRISEs* in the restatement episode window. The consensus forecast for the first *ESURPRISE* after the restatement announcement is taken immediately after the restatement announcement. The consensus forecasts for subsequent *ESURPRISEs* are taken immediately after the previous *ESURPRISE*. Winsorized at 1 and 99 percent.
FIRMSIZE	Natural log of the firm's market capitalization two days before the restatement announcement.
FRAUD	Equals 1 if the errors are described as intentional or if investigations by a government entity, the board of directors, or audit committee are made public (Hennes et al. 2008). Equals 0 otherwise or if the fraud was committed in lower levels of the organization.
FRAUD_LOWLEVEL	Equals 1 if a fraud or investigation pertains to errors committed in subsidiaries or lower levels of the organization, 0 otherwise.
LONG_MAG_DELAY	Equals 1 if more than 30 days elapse between the initial restatement announcement and the announcement of the restatement's definitive earnings impact, 0 otherwise.
MAG	Cumulative impact of the restatement on past earnings, scaled by total assets for the fiscal year ended prior to the restatement announcement. Winsorized at 1 and 99 percent.
MBE	Equals 1 if earnings are announced in the three-day window around the restatement announcement and they meet or beat the most recent consensus analyst forecast on I/B/E/S, 0 otherwise.

(continued on next page)

Panel C: Other Explanatory Variables

MGT_PROMPTS	Equals 1 if the GAO report identifies the company as the prompter of the restatement (as opposed to the auditor or the SEC), 0 otherwise.
MULTIPLE	Equals 1 if the restatement spans item categories or involves three or more errors in the same category, 0 otherwise. The categories are (1) CORE income items, (2) depreciation, amortization, equity-based compensation, and other non-cash components of pre-tax operating income, (3) three common operating lease accounting errors announced by many firms in 2004 and 2005 (Acito et al. 2009), (4) taxes, (5) derivatives, (6) impairments and other valuation issues associated with noncurrent assets and liabilities, (7) off-balance-sheet obligations, (8) financing activities, (9) balance sheet or statement of cash flow reclassifications not affecting income, (10) merger-related or special items not included in other categories, and (11) unspecified or other errors.
NO_INITIAL_MAG	Equals 1 if no definitive earnings impact is disclosed when the initial restatement announcement is made, 0 otherwise.
POSITIVE_MAG	Equals 1 if cumulative restated earnings are greater than or equal to cumulative original earnings, 0 otherwise.
POSITIVE_RETURN	Equals 1 if *RETURN_ANNOUNCE* > 0, 0 otherwise.
QUARTERLY_ONLY	Equals 1 if the restatement involves only prior quarters of the current fiscal year, 0 otherwise.
RETURN_PRIOR	Size-adjusted buy-and-hold stock return over days $(-180, -2)$ relative to the initial restatement announcement. Winsorized at 1 and 99 percent.
VOLATILITY_INDEX	Mean value of the Chicago Board Options Exchange's volatility index over the three-day or episode return windows.

REFERENCES

Accounting Principles Board (APB). 1971. *Accounting Changes*. Accounting Principles Board Opinion No. 20. Norwalk, CT: APB.

Acito, A., J. J. Burks, and W. B. Johnson. 2009. Materiality decisions and the correction of accounting errors. *The Accounting Review* 84 (3): 659–688.

Advisory Committee on Improvements to Financial Reporting (CIFR). 2008. Final report of the Advisory Committee on Improvements to Financial Reporting to the United States Securities and Exchange Commission. Available at: http://www.sec.gov/about/offices/oca/acifr/acifr-finalreport.pdf.

Altamuro, J., A. L. Beatty, and J. Weber. 2005. The effects of accelerated revenue recognition on earnings management and earnings informativeness: Evidence from SEC staff accounting bulletin no. 101. *The Accounting Review* 80 (2): 373–401.

Badertscher, B., and J. J. Burks. 2010. Accounting restatements and the timeliness of disclosures. Working paper, University of Notre Dame.

Bailey, W., H. Li, C. X. Mao, and R. Zhong. 2003. Regulation fair disclosure and earnings information: Market, analyst, and corporate responses. *The Journal of Finance* 58 (6): 2487–2514.

Bamber, L. S., and Y. S. Cheon. 1995. Differential price and volume reactions to accounting earnings announcements. *The Accounting Review* 70 (3): 417–441.

———, O. E. Barron, and T. L. Stober. 1997. Trading volume and different aspects of disagreement coincident with earnings announcements. *The Accounting Review* 72 (4): 575–597.

Beaver, W. H. 1968. The information content of annual earnings announcements. *Journal of Accounting Research* 6 (Supplemental Issue): 67–92.

———, M. McNichols, and R. Price. 2007. Delisting returns and their effect on accounting-based market anomalies. *Journal of Accounting and Economics* 43 (2–3): 341–368.

Belsley, D. A., K. Kuh, and R. E. Welsch. 1980. *Regression Diagnostics*. New York, NY: John Wiley & Sons, Inc.

Black, F. 1986. Noise. *The Journal of Finance* 41 (3): 529–543.

Bloomfield, R. J. 2002. The incomplete revelation hypothesis and financial reporting. *Accounting Horizons* 16 (3): 233–243.

Burks, J. J. 2010. Disciplinary measures in response to restatements after Sarbanes-Oxley. *Journal of Accounting and Public Policy* 29 (3): 195–225.

Byun, J., and M. S. Rozeff. 2003. Long-run performance after stock splits: 1927 to 1996. *The Journal of Finance* 58 (3): 1063–1086.

Chartered Financial Analyst (CFA) Institute. 2008. Letter to Ms. Nancy M. Morris, Federal Advisory Committee Management Officer, U.S. Securities and Exchange Commission. (March 31). Available at: http://www.sec.gov/comments/265-24/26524-68.pdf.

Consumer Federation of America. 2008. Letter to Ms. Nancy Morris M. Morris [sic], Federal Advisory Committee Management Officer, U.S. Securities and Exchange Commission. (April 14). Available at: http://www.sec.gov/comments/265-24/26524-84.pdf.

Cummings, J. 2008. Toward betting reporting: A talk with Robert Pozen. *BusinessFinanceMag.com* (Feb 12). Available at: http://businessfinancemag.com/article/toward-better-reporting-talk-robert-pozen-0212.

Daniel, K., D. Hirshleifer, and A. Subrahmanyam. 1998. Investor psychology and security market under- and over-reactions. *The Journal of Finance* 53 (6): 1839–1885.

Der, G., and B. S. Everitt. 2002. *A Handbook of Statistical Analysis Using SAS*. 2nd edition. Boca Raton, FL: Chapman & Hall/CRC.

D'Souza, J., K. Ramesh, and M. Shen. 2010. The interdependence between institutional ownership and information dissemination by data aggregators. *The Accounting Review* 85 (1): 159–193.

Fama, E. F., and K. R. French. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33 (1): 3–56.

Files, R., E. P. Swanson, and S. Tse. 2009. Stealth disclosure of accounting restatements. *The Accounting Review* 84 (5): 1495–1520.

Financial Accounting Standards Board (FASB). 2005. *Accounting Changes and Error Corrections*. Statement of Financial Accounting Standards No. 154. Norwalk, CT. FASB.

Garfinkel, J. A., and J. Sokobin. 2006. Volume, opinion divergence, and returns: A study of post-earnings announcement drift. *Journal of Accounting Research* 44 (1): 85–112.

Gleason, C. A., N. T. Jenkins, and W. B. Johnson. 2008. The contagion effects of financial statements. *The Accounting Review* 83 (1): 83–110.

Gordon, E. A., E. Henry, M. Peytcheva, and L. Sun. 2008. Disclosure credibility and market reaction to restatements. Working paper, Rutgers University.

Grossman, S. J., and J. E. Stiglitz. 1980. On the impossibility of informationally efficient markets. *The American Economic Review* 70 (3): 393–408.

Heffes, E. M. 2007. CIFR chair Robert Pozen: Committee seeks more than just to reduce complexity. *Financial Executive* 23 (9): 16–18.

Hennes, K. M., A. J. Leone, and B. P. Miller. 2008. The importance of distinguishing errors from irregularities in restatement research: The case of restatements and CEO/CFO turnover. *The Accounting Review* 83 (6): 1487–1519.

Hirschey, M., and K. R. Smith. 2009. Valuation effects of accounting restatements in the post-SOX environment. Working paper, University of Kansas.

Hirshleifer, D., and S. H. Teoh. 2003. Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics* 36 (1–3): 337–386.

Hranaiova, J., and S. L. Byers. 2007. Changes in market responses to financial statement restatement announcements in the Sarbanes-Oxley era. Working paper, Public Company Accounting Oversight Board.

Huang, Y. 2010. When and why did the market reactions to restatement announcements attenuate? Working paper, University of Kansas.

Ikenberry, D., J. Lakonishok, and T. Vermaelen. 1995. Market under-reaction to open market share repurchases. *Journal of Financial Economics* 39: 181–208.

Johnson, S. 2008. Proposed restatement guidelines draw investor alarm. *CFO.com* (March 17). Available at: http://careers.cfo.com/article.cfm/10869929?f=bc.

Kandel, E., and N. D. Pearson. 1995. Differential interpretation of public signals and trade in speculative markets. *The Journal of Political Economy* 103 (4): 831–872.

Karpoff, J. M. 1987. The relation between price changes and trading volume: A survey. *Journal of Financial and Quantitative Analysis* 22 (1): 109–126.

Kim, O., and R. E. Verrecchia. 1997. Pre-announcement and event-period private information. *Journal of Accounting and Economics* 24 (3): 395–419.

Lee, I. 1997. Do firms knowingly sell overvalued equity? *The Journal of Finance* 52 (4): 1439–1466.

Lev, B. 2003. Corporate earnings: Facts and fiction. *The Journal of Economic Perspectives* 17 (2): 27–50.

———, S. G. Ryan, and M. Wu. 2008. Rewriting earnings history. *Review of Accounting Studies* 13 (4): 419–451.

Lipe, R. C., L. Bryant, and S. K. Widener. 1998. Do nonlinearity, firm-specific coefficients, and losses represent distinct factors in the relation between stock returns and accounting earnings? *Journal of Accounting and Economics* 25 (2): 195–214.

Lundholm, R. 1999. Reporting on the past: A new approach to improving accounting today. *Accounting Horizons* 13 (4): 315–323.

Lyon, J. D., B. M. Barber, and C. Tsai. 1999. Improved methods for tests of long-run abnormal stock returns. *The Journal of Finance* 54 (1): 165–201.

Myers, L. A., S. Scholz, and N. Y. Sharp. 2010. Restating under the radar? Determinants of restatement disclosure choices and the related market reactions. Working paper, University of Arkansas.

Palmrose, Z.-V., V. J. Richardson, and S. Scholz. 2004a. Determinants of market reactions to restatement announcements. *Journal of Accounting and Economics* 37 (1): 59–89.

———, and S. Scholz. 2004b. The accounting causes and legal consequences of non-GAAP reporting: Evidence from restatements. *Contemporary Accounting Research* 21 (1): 139–180.

Paulson, H. 2007. The key test of accurate financial reporting is trust. *Financial Times* (May 17): 15.

Plumlee, M., and T. L. Yohn. 2010. An analysis of the underlying causes attributed to restatements. *Accounting Horizons* 24 (1): 41–64.

Rummell, N. 2008. Tumble in restatements sparks criticism of SEC. *Financial Week* (August 25). Available at: http://www.financialweek.com/apps/pbcs.dll/article?AID=/20080825/REG/860815.

Scholz, S. 2008. The changing nature and consequences of public company financial restatements. The U. S. Department of the Treasury. Available at: http://www.treasury.gov/press-center/press-releases/Documents/FinancialRestatements_1997_2006.pdf.

Securities and Exchange Commission. 2002. Frequently asked questions about Commission Order No. 4-460. Available at: http://www.sec.gov/rules/other/4-460faqs.htm.

Shleifer, A., and R. W. Vishny. 1997. The limits of arbitrage. *The Journal of Finance* 52 (1): 35–55.

Stuart, A. 2010. Restatements on the decline. *CFO.com* (March 4). Available at: http://www.cfo.com/article.cfm/14480266.

Taub, S. 2009. The state of restatements: Sharply falling. *CFO.com* (March 10). Available at: http://www.cfo.com/article.cfm/13270658?f=search.

U.S. Government Accountability Office (GAO). 2002. *Financial Statement Restatements: Trends, Market Impacts, Regulatory Responses, and Remaining Challenges*. GAO-03-138. Washington, D.C.: GAO.

———. 2006. *Financial Restatements: Update of Public Company Trends, Market Impacts, and Regulatory Enforcement Activities*. GAO-06-678. Washington, D.C.: GAO.

———. 2007. *Financial Restatement Database*. GAO-06-1053R. Washington, D.C.: GAO.

U.S. House of Representatives. 2002. *The Sarbanes-Oxley Act of 2002*. Public Law 107-204 Washington, D.C.: Government Printing Office.

White, H. 1980. A heteroscedastic-consistent covariance matrix estimator and a direct test of heteroscedasticity. *Econometrica* 48 (4): 817–838.

Wright, A., C. Botosan, R. Colson, M. Eining, P. Hopkins, D. Plumlee, M. Plumlee, J. Rebele, P. Reckers, S. Scholz, K. St. Pierre, and L. Vincent. 2008. Response to the progress report of the SEC Advisory Committee on Improvements to Financial Reporting. *Accounting Horizons* 22 (4): 471–494.

THE ACCOUNTING REVIEW
Vol. 86, No. 2
2011
pp. 541–568

American Accounting Association
DOI: 10.2308/accr.00000020

Asset Securitization, Securitization Recourse, and Information Uncertainty

Mei Cheng
The University of Arizona

Dan S. Dhaliwal
The University of Arizona and Korea University Business School

Monica Neamtiu
The University of Arizona

ABSTRACT: In this study, we examine some of the consequences of asset securitization. Specifically, using a sample of bank holding companies, we investigate whether the difficulty in assessing the true extent of risk transfer, between securitizing banks and investors in asset-backed securities, affects bank information uncertainty. We find that when market participants have a greater difficulty in estimating risk transfer, banks face greater information uncertainty (i.e., larger bid-ask spreads and analyst forecast dispersion). In addition, we find that this effect is mitigated for banks that operate in a higher quality information environment. We also find that banks that securitize financial assets have higher spreads and analyst forecast dispersion as compared to non-securitizing banks.

Keywords: *securitization; recourse; information uncertainty; banks.*

JEL Classifications: *G14; G21.*

I. INTRODUCTION

The volume of securitized financial assets has increased considerably over the last two decades. According to the Bond Market Association, as of the end of the first quarter of 2007, the asset-backed market comprised roughly \$8.9 trillion in outstanding securities compared to \$5.4 trillion in outstanding corporate bonds and about \$4.5 trillion in outstanding U.S. Treasury debt.

The rapid growth in securitization activities and the recent turmoil affecting securitization markets raise interesting questions regarding the benefits and costs of this type of financing

We thank two anonymous referees, Stephen Ryan, Wayne Thomas (editor), Mark Trombley, and the seminar participants at the University of Southern California, University of Auckland, Vanderbilt University, London School of Business, National Taiwan University, The Pennsylvania State University 2009 Research Conference, the FARS 2009 Meeting, and the CAPANA 2009 Conference for their valuable comments and suggestions.

Editor's note: Accepted by Wayne Thomas.

Submitted: November 2008
Accepted: July 2010
Published Online: March 2011

mechanism. Prior academic and practitioner literature focuses primarily on the incentives and benefits of securitizing financial assets (e.g., Ryan 2007; Clarkson et al. 2002). This study extends prior research by examining some of the consequences of asset securitization. Specifically, we investigate the extent to which the difficulty in assessing the extent of risk transfer from securitizing banks to investors in asset-backed securities is associated with measures of bank information uncertainty.[1] Assessing the true extent of risk transfer is a critical issue in the securitization context and a significant focus of both academic research and standard-setters (Schipper and Yohn 2007).

We hypothesize that the difficulty in assessing the extent of recourse is positively associated with securitizing banks' information uncertainty. We expect market participants to have difficulty assessing the true extent of securitization recourse because of the complexity and lack of transparency associated with asset securitization transactions.[2] In practice, securitization transactions can have very complex structures. Banks typically rely on multiple special purpose entities (SPEs) to structure these transactions. In addition, various securitization structures can differ substantially in the extent to which issuers retain the risks associated with the securitized assets, ranging from a minimum to almost complete risk transfer. More importantly, investors and financial analysts have repeatedly indicated that the amount of securitization disclosure is insufficient and greater transparency is needed to fully understand the extent of risk transfer associated with securitized financial assets (FASB 2008; CFA Institute 2008). The difficulty in estimating risk transfer is likely to lead to information asymmetry among market participants if some market participants have better information and/or better information-processing abilities (Barth et al. 2003). It is also possible that the difficulty in assessing securitization risk transfer generates disagreement even among investors with similar levels of expertise and sophistication (Schwarcz 2004).

Because market participants' uncertainty with respect to the extent of risk transfer is not directly observable, in our empirical analyses we select several measures we expect to be correlated with the uncertainty about the extent of recourse. We use banks' securitization volume, securitization income, the proportion of nonperforming loans, charge-offs associated with securitized loans, and the volume of asset-backed securities retained by securitizing banks on their balance sheets. To mitigate the concern that each of these individual proxies measures the underlying concept of interest (i.e., the uncertainty about the extent of risk transfer) with error, we use principal component analysis and construct a recourse uncertainty factor based on these measures. We test our hypothesis by regressing proxies for bank information uncertainty, such as bid-ask spreads and earnings forecast dispersion, on the recourse uncertainty factor and control variables for a sample of securitizing banks. Our tests consistently document a positive association between the recourse uncertainty factor and banks' bid-ask spreads and analyst forecast dispersion.

In light of these findings, we also investigate whether the quality of the information environment (i.e., the quality of relevant information and the degree of investor sophistication) impacts the relation between the difficulty in estimating recourse and a securitizing bank's information uncertainty. Cross-sectional tests about how the effect of recourse uncertainty varies with the

[1] In the context of this study, by "greater information uncertainty" we mean lower precision of financial information and/or greater information asymmetry among information users. The information uncertainty is equivalent to the estimation risk (i.e., uncertainty in the parameters of the model or parameter uncertainty) discussed in the prior literature (Klein and Bawa 1976; Barry and Brown 1985). Empirically, it is the residual uncertainty after controlling for conventional risk measures. Prior literature (e.g., Barry and Brown 1985; Clarkson et al. 1996; Easley and O'Hara 2004; Amihud and Mendelson 1986, 1989) shows that greater information uncertainty can have real economic costs in the form of a higher cost of capital.

[2] In this study, we broadly refer to risk retained through both contractual obligations (i.e., retained interests) and noncontractual promises (i.e., implicit support) as recourse. For a detailed discussion about the importance of recourse in securitization transactions, see the "Background and Related Literature" and "Hypotheses Development" sections.

quality of the information environment can also help mitigate the concern that, despite controlling for risk measures frequently used in prior literature, our recourse uncertainty proxy simply captures the effect of securitization on bank intrinsic risk and not the effect on bank information uncertainty. If our recourse uncertainty measure captures the information effect of securitization, then we hypothesize that the impact of the difficulty in estimating recourse on bank information uncertainty will be mitigated for banks that operate in a higher quality information environment.

When we include in our main regressions interaction terms between the recourse uncertainty measure and the quality of the information environment proxies (i.e., analyst following, institutional investor holdings, and bank size), the findings support our hypothesis that the effect of recourse on banks' bid-ask spread and forecast dispersion is mitigated for banks with larger analyst following, larger institutional investor holdings, and greater size. In addition, when we interact the recourse uncertainty measure with information environment and intrinsic risk proxies, our information environment results continue to hold, while the impact of intrinsic risk on the relation between recourse uncertainty and bid-ask spreads/forecast dispersion is not statistically significant.

Given our finding that more difficulty in assessing risk transfer for securitizing banks is positively associated with greater bank information uncertainty, an interesting related question is whether securitization results in banks that engage in this type of financing transactions being more or less exposed to information uncertainty, as compared to banks that do not securitize assets. Securitization transactions can impact information uncertainty through several mechanisms. On the one hand, compared to non-securitizing banks, securitizing banks are exposed to uncertainty about the true level of risk transfer associated with the transferred assets and uncertainty about adequate reinvestment opportunities for the large cash amounts generated through securitization. On the other hand, securitization could lead to lower information uncertainty for securitizing banks relative to non-securitizing banks because of additional disclosure requirements for securitized assets.[3] We find that, after controlling for other determinants of information uncertainty, banks that securitize financial assets have higher spreads and analyst forecast dispersion as compared to banks that do not engage in securitization. These results are robust to controlling for potential endogeneity bias.

This study contributes to the literature by investigating the effect of the difficulty in assessing recourse (a key securitization feature) of bank information uncertainty. Many of the prior securitization studies (e.g., Niu and Richardson 2006; Landsman et al. 2008; Chen et al. 2008) examine whether, in their assessments of firm systematic risk, investors treat securitizations as a form of secured borrowing. They find that, *on average*, market participants view off-balance sheet securitized assets as having (to an extent) similar risk and value relevance as the assets kept on the issuer's balance sheet. In contrast, we follow the suggestion for future research issued by Schipper and Yohn (2007) to "provide additional insight into other investor judgments and decisions" related to securitization. Our study focuses on how market participants *differ* in their understanding and estimation of the degree of risk transfer (i.e., how market participants' difficulty in assessing the level of risk transfer affects the information uncertainty of securitizing banks).

In addition, our findings suggest that during our sample period, at least for some banks, the required disclosures did not provide enough information for market participants to sufficiently understand the level of securitization recourse. Our results also indicate that the impact of uncer-

[3] When issuing asset-backed securities, securitizing banks are required to disclose more information about these assets (through the process of registering the asset-backed securities with the SEC) compared to non-securitizing banks that keep the loans on the balance sheet (Schwarcz 2004; Foley et al. 1999). The additional required disclosure can lead to reduced information uncertainty with respect to not only the securitized assets, but also to the securitizing firms in general (Foley et al. 1999; Berger and Udell 1995; Cantor and Rouyer 2000).

tainty about the level of risk transfer on bank information uncertainty can be mitigated for securitizing banks that operate in a high-quality information environment. Therefore, this study should be of interest to accounting regulators because it highlights the importance of high-quality disclosure in understanding the impact of securitization risk transfer. These findings support the FASB's concerns that companies provided insufficient and hard-to-interpret securitization disclosures (FASB 2002, 2008) and the FASB's recent efforts to improve such disclosures. Future research can examine whether the recently enhanced disclosures for securitized assets (ASC 860-10, formerly SFAS No. 166, FASB 2009a, 2009b) mitigate the impact of securitization recourse on firms' information uncertainty.

The next section provides background information. Section III develops our hypotheses. Section IV presents our sample selection and descriptive statistics. Section V discusses the research design and the findings. Section VI concludes the study.

II. BACKGROUND AND RELATED LITERATURE

Accounting Treatment

Under ASC 860 (formerly SFAS No. 140, FASB 2000), securitizations can be accounted for as either sales or secured borrowings. For a transfer of financial assets to an SPE to qualify for sale accounting treatment, it must meet the following criteria: (1) the assets are isolated from the transferor and its creditors even in bankruptcy; (2) the SPE has the right to pledge or exchange the assets; and (3) the transferor does not maintain effective control over the assets through certain forms of continuing involvement.[4] If the securitization receives sale accounting treatment, then the transferor: (1) removes the assets from its balance sheet; (2) records cash proceeds in the amount received and recognizes any noncash proceeds in the securitized assets at fair value; (3) recognizes retained asset-backed securities at the book value of the securitized assets times the fair value of the retained securities divided by the fair value of the securitized assets; (4) recognizes retained contractual interests other than asset-backed securities, (e.g., servicing assets and recourse liabilities) at fair value; and (5) records a gain or loss on the sale to balance the journal entry. If the securitization is accounted for as a sale and additional conditions are satisfied (see Gorton and Souleles 2006), then the issuer does not have to consolidate the assets and liabilities held by its SPEs. If the transfer is accounted for as secured borrowing, then the financial assets remain on the balance sheet and the issuer recognizes a liability for the proceeds from the transfer.

Securitization Recourse

In a securitization transaction, the securitizing bank transfers pools of assets (such as mortgages or trade and credit card receivables) to SPEs. The SPEs finance these assets by selling to investors various classes of securities backed by the transferred asset pools and the cash flows they generate. The securities issued in this manner are commonly referred to as asset-backed securities. Investors in asset-backed securities face information asymmetry stemming from securitizing banks having better information about the transferred assets than do they. To mitigate potential adverse-selection problems, the securitizing bank offers to the outside investors some form of recourse. In the context of securitization transactions, recourse usually referrers to guaranties promised to

---

[4] In practice, banks typically structure securitization transactions by transferring pools of assets to qualified special purpose entities (QSPEs). QSPEs are exempt from consolidation requirements under ASC 860 (formerly SFAS No. 140), the accounting standard that covers securitization transactions during our sample period. In 2009, the FASB issued two new accounting standards, ASC 860-10 and ASC 810 (formerly SFAS No. 166 and SFAS No. 167, respectively) that determine whether securitizations and other transfers of financial instruments are given off-balance sheet treatment. The revisions to the securitization standards will result in many existing off-balance sheet securitizations being treated as secured financing and added back on-balance sheet.

investors in asset-backed securities allowing the transfer of some losses back to the originating bank if the performance of the underlying pools of securitized assets deteriorates.

Recourse can take an explicit form. For example, in many securitization transactions, the securitizing banks retain on their balance sheet the most subordinated asset-backed securities (often referred to as retained interests) that bear the first loss from the pool of securitized assets. The banks sell to outside investors the senior asset-backed securities.

To maintain their reputation and future access to securitization markets, issuers can also offer an implicit (i.e., non-contractual) promise to support underperforming securitized assets. Loss of reputation (i.e., the ability to sell loans economically) can expose banks to decreased liquidity, increased interest rate risk, and potential burdensome regulatory supervision. An originating bank can provide implicit support in various ways. In 2002, the Federal Financial Institutions Examination Council (FFIEC) released a document to assist in identifying cases of implicit recourse. The guidance lists four major actions that signal possible implicit recourse: (1) selling assets to a securitization trust or other SPE at a discount from the price specified in the securitization documents, which is typically par value; (2) purchasing assets from a trust or other SPE at an amount greater than fair value; (3) exchanging performing assets for non-performing assets in a trust or other SPE; and (4) funding credit enhancements beyond contractual requirements.[5]

Anecdotal evidence and academic research (Calomiris and Mason 2004; Higgins and Mason 2004; Gorton and Souleles 2006; Vermilyea et al. 2008) indicate that many issuers have chosen to implicitly support troubled transactions rather than impose credit losses on investors in asset-backed securities. The academic literature on implicit recourse focuses both on reasons for its existence and the effects. Gorton and Souleles (2006) provide theoretical arguments that the existence of implicit recourse is meant to reduce moral hazard problems. In the context of credit card securitizations, Calomiris and Mason (2004) find that this implicit recourse is more in line with an optimal contracting view of securitization rather than an attempt to reduce regulatory capital requirements and avoid early amortization. Prior literature also documents that recourse provides some benefits to the originating banks, including improved short- and long-term stock price performance and financial performance (e.g., Higgins and Mason 2004). In contrast, our study emphasizes the consequences for the originating banks stemming from market participants' difficulty in assessing the true level of securitization recourse.

Recent literature (e.g., Mason 2009) discusses the sources of information asymmetry that are related to recourse, including not knowing the distribution of external shocks among various classes of securitized assets, the lack of understanding about the triggers and terms of securitization recourse, and the difficulty in estimating the impact of recourse. Mason (2008) even suggests that, given the large amount of asymmetric information in poor securitization disclosure, it is not surprising that a crisis occurred once securitization reached a sufficient scale. Our study extends this line of literature by empirically measuring recourse uncertainty and examining its information consequences at the securitizing bank level.

III. HYPOTHESES DEVELOPMENT

Because of the possibility of recourse, a critical issue in understanding the impact of securitizing assets on banks that originate such transactions is the extent to which the credit risk

---

[5] These four FFEIC actions can signal implicit recourse, but their presence alone is not sufficient to definitely conclude the existence of implicit recourse. Regulators that want to prove a failure to consolidate in the presence of implicit recourse would have to further investigate each particular case and its specific details. In addition, implicit recourse can also take forms that are more subtle than the four FFEIC events listed above. For example, anecdotal evidence indicates that banks sometimes misclassify certain credit losses as fraud losses to hide the fact that they did offer implicit recourse.

associated with the securitized assets has been transferred from the originating bank to those who invest in asset-backed securities. To understand the effect of recourse on the securitizing banks, market participants have to estimate: (1) the anticipated need for future support for troubled securitization transactions (i.e., expected future losses associated with securitized assets) and (2) banks' willingness and potential methods of offering support, if the need arises.

Market participants are likely to find it difficult to assess the true extent of securitization recourse because securitization transactions are very complex. In practice, various features of securitization transactions (such as the magnitude of retained asset-backed securities, the level of subordination of retained interests, the credit enhancement offered) and thus, the extent of risk transfer, vary considerably even across securitization transactions originated by the same bank. As for implicit recourse obligations, information users find it difficult to even identify (let alone measure) such obligations, because banks may or may not support their troubled securitizations at their discretion.

Note that the complexity and heterogeneity of securitization transactions would not lead to uncertainty with respect to the true extent of securitization risk transfer if publicly available securitization-related disclosure is adequate or market participants are able to use their private sources of information to clearly assess the extent of risk transfer. However, there are reasons to believe that the quality of securitization-related disclosure is inadequate. Financial statement users repeatedly indicate that the amount of securitization disclosure is insufficient and greater transparency is needed to fully understand the extent of risk transfer associated with securitized financial assets (FASB 2008). Further, financial analysts surveyed by the CFA Institute in 1999, 2003, and 2007 consistently rated securitization-related disclosure as high in importance, yet assigned low ratings for the quality of such disclosure (CFA Institute 2008).

In the securitization context, the difficulty in estimating risk transfer can lead to information asymmetry and disagreement among market participants if some market participants have better information and/or better information-processing abilities. Prior studies (Ryan 2007; Schwarcz 2004) argue that financial reporting requirements simplify considerably the underlying economics behind asset securitizations and, therefore, cannot fully describe these complex transactions. Although the risk transfer associated with securitized assets varies on a continuum, the financial reporting of securitizations offers only a simplified dichotomous choice (i.e., sale versus secured borrowing accounting).[6] In practice, the majority of securitization transactions are recorded using sale accounting and the transactions are kept off-balance sheet. Because of this financial reporting treatment, information users have to rely on footnotes to the financial statements and other disclosure sources. Prior literature (e.g., Barth et al. 2003) suggests that reliance on disclosed rather than recognized information leads to information asymmetry among market participants, insofar as investors with a higher level of expertise are able to collect information from disclosures at a lower cost and/or a faster pace, thus obtaining an informational advantage over investors with less expertise. Mason (2009) discusses some specific sources of information asymmetry related to securitization recourse, such as not knowing the distribution of external shocks among various classes of securitized assets, lack of understanding about the triggers and terms of securitization recourse, and difficulty in estimating the impact of recourse on the originating bank. In addition, difficulty in assessing the extent of recourse can generate disagreement even among investors with similar levels of expertise and sophistication. Some authors (e.g., Schwarcz 2004) suggest that securitization transactions are so complex that the sophistication of their disclosure provides a

[6] Under U.S. GAAP, sale accounting is permitted when a firm surrenders control over the securitized assets (see Section II for a more detailed discussion of this point).

level of sophistication that is challenging even for institutional investors and security analysts. Therefore, we hypothesize that measures of difficulty in assessing the extent of recourse are positively associated with securitizing banks' information uncertainty.[7]

To be sure, some banks have a higher quality information environment (i.e., a higher quality of relevant information and degree of investor sophistication). For securitizing banks that operate in a higher quality information environment, market participants have better information about investment opportunities, lending standards, alternative sources of financing, and bank reputation concerns. This type of information is critical in anticipating a bank's need and willingness to provide future support for its troubled securitization transactions. In addition, for securitizing banks that have a greater percentage of institutional ownership, investors have a higher level of expertise on which to rely when collecting and processing information about complex securitization transactions. Therefore, we hypothesize that the impact of the difficulty in estimating recourse on bank information uncertainty will be mitigated for banks that operate in a higher quality information environment. These cross-sectional predictions about how the effect of recourse uncertainty varies with the quality of the information environment can also help mitigate the concern that, despite controlling for risk measures frequently used in prior literature, our recourse uncertainty analysis simply captures the effect of securitization on bank intrinsic risk and not the effect on bank information uncertainty.[8]

IV. SAMPLE SELECTION AND DESCRIPTIVE STATISTICS

In testing our hypotheses, we focus on the impact of the difficulty in estimating securitization recourse on the level of information uncertainty in securitizing banks. While other types of firms securitize financial assets as well, we choose to investigate the loan securitizations of banks for several reasons. First, given that banks operate in a regulated industry, their financial reporting of securitizations is substantially more homogenous than the reporting of firms in various other industries. Therefore, focusing on bank securitizations allows us to reduce the impact of variations in the regulatory disclosure level on bank information uncertainty. Second, by focusing on the banking industry, we mitigate data availability concerns. In contrast to banks, firms in other industries vary considerably more in the degree to which they present comprehensive securitization information. Third, banks are major participants in securitization markets. Prior studies (e.g., Dechow et al. 2010) that use a multi-industry sample find that only 30 percent of their sample firms come from nonfinancial industries and banks are the largest category among the financial firms engaged in securitizations.

We start our sample selection by identifying all bank holding companies (hereafter, banks) with the necessary quarterly financial information available from the regulatory Y-9C reports that

---

7 However, one may conceive that, for some banks, market participants have no difficulty in estimating the extent of securitization recourse. For example, it is possible that for some banks every market participant agrees that the bank's securitization transactions are true sales of assets or secured borrowings. This possibility adds tension to our hypothesis.

8 From a conceptual point of view, there is an important distinction between intrinsic risk and information uncertainty. Following recent literature (e.g., Anderson et al. 2009), we adopt the position that an event is *risky* if its outcome is unknown, but the distribution of its potential outcomes is known; an event is *uncertain* if its outcome is unknown and the distribution of its outcomes is also unknown. To illustrate this difference in our context we offer an example. For a bank that securitizes assets, market participants know there is a possibility of future recourse. To compute the expected value of future support, market participants take into account a set of possible future economic outcomes and attach a point probability estimate to each outcome. This scenario assumes that market participants are able to develop a *point* probability for each possible outcome. In this case, the market participants face risk with respect to securitization recourse. However, in some circumstances, market participants do not have accurate and complete information to assign a point probability (i.e., the distribution of probabilities for each outcome cannot be reduced to a single point). In this case, market participants face information uncertainty with respect to securitization recourse. Our concept of information uncertainty is similar to the notion of Knightian uncertainty, long established in the economic literature (e.g., Keynes 1937; Knight 1921).

any bank with consolidated assets of at least $150 million ($500 million after March 2006) must file with the Federal Reserve.[9] We eliminate banks whose common equity is not traded on the NYSE, NASDAQ, or AMEX stock exchanges, as one of our information uncertainty measures (i.e., the effective bid-ask spread) is market-based. From the Y-9C reports, we gather bank-quarter information from 2001 (second calendar quarter) to 2007 (second calendar quarter). We identify banks that securitize assets and collect information about our securitization variables from schedule HC-S (Servicing, Securitization, and Asset Sale Activities) of the Y-9C reports.[10] Our sample starts with the second quarter of 2001 because the Schedule HC-S was first available then. We end our sample period with the second quarter of 2007 because we are interested in studying the impact of securitization transactions on information uncertainty under normal market conditions (i.e., before the current securitization crisis). The securitization transactions reported on the Schedule HC-S are accounted for as sales. Consistent with prior research (e.g., Chen et al. 2008; Landsman et al. 2008), we focus on securitizations accounted for as sales, because typically the characteristics of securitizations accounted for as secured borrowings are unobservable to bank outsiders.

Following prior studies (Chordia et al. 2001; Flannery et al. 2004), we use both the consolidated trades and the consolidated quote files from the TAQ database to compute the effective bid-ask spreads. We use CRSP information to compute the share turnover and return volatility measures, and the I/B/E/S summary file to compute analyst forecast dispersion and analyst following measures. After eliminating observations with missing values for the necessary control variables used in each of our empirical models, we obtain a sample of 871 and 687 bank-quarter securitization observations for the spread and dispersion models, respectively.

Table 1 presents descriptive statistics for the more comprehensive sample used in our bid-ask spread analysis (n = 871 bank-quarter observations). For the banks in our sample, net loans represent 60.96 percent of total assets, quarterly net income is 0.66 percent of total assets, the liability-to-equity ratio averages about 11.07, and the log value of total assets is 9.88. Bank return volatility has a mean of 0.015 and the average price per share is $34.78. Close to a third of our sample banks (31.34 percent) are listed on the NASDAQ, and the rest are listed on the NYSE or on the AMEX. In general, the descriptive statistics for our sample are comparable to those reported by Flannery et al. (2004) and Chen et al. (2008). For the sample used in the forecast dispersion analysis (n = 687), the untabulated mean (median) value of forecast dispersion is 0.021 (0.013). All other descriptive statistics are similar to the statistics presented in Table 1.

V. RESEARCH DESIGN AND FINDINGS

Information Uncertainty Measures

To test our hypothesis, we use two proxies for bank information uncertainty: bid-ask spreads and analyst forecast dispersion. The bid-ask spread is commonly used in the literature as a measure of information asymmetry (e.g., Flannery et al. 2004; Leuz and Verrecchia 2000).[11] When

---

[9] A bank holding company is any firm that controls a bank, where control is defined as holding more than 25 percent of a bank's equity or having the ability to elect more than two directors. Our sample of bank holding companies does not include investment banks.

[10] We include in our sample all the available quarters for the securitizing banks, even if a bank does not securitize in each quarter during our sample period. To test the sensitivity of our results to this sample selection choice, we run two types of robustness tests. We rerun our analyses using: (1) only those bank-quarter observations with positive securitization volume and (2) only banks that have at least six quarters with positive securitization volume during our sample period (following Chen et al. 2008). Our results do not change.

[11] We note that, besides an adverse-selection component, the bid-ask spreads also include an order-processing component and an inventory-holding component. In this study, we are interested in the adverse-selection component of the spreads. However, the empirical decomposition of the bid-ask spreads into the three distinct components is not very reliable

TABLE 1

Distributional Characteristics of the Variables Used in the Multivariate Analysis (n = 871)

Variable	Mean	Median	Q1	Q3	Std. Dev.
SPREAD	0.0056	0.0023	0.0013	0.0059	0.0080
NETLNS	0.6096	0.6544	0.5400	0.7136	0.1569
SIZE	9.8841	10.3482	8.5890	11.4642	2.1311
PROFIT	0.0066	0.0062	0.0032	0.0099	0.0052
MVLEV	11.0740	10.3285	9.3194	12.2134	3.0155
STDRET	0.0151	0.0133	0.0104	0.0181	0.0070
PRICE	34.7803	30.5400	23.1700	42.3200	17.5294
NASDAQ	0.3134	0.0000	0.0000	1.0000	0.4642
TURNOVER	0.1118	0.0881	0.0586	0.1387	0.0868
FOLLOW	2.1087	2.6391	1.3863	2.9957	1.1484
DERIVATIVE	1.3407	0.1378	0.0234	0.6977	3.4736
CHGOFF_ONBS	0.0034	0.0022	0.0009	0.0043	0.0038
NPL_ONBS	0.0112	0.0100	0.0060	0.0144	0.0074
GAP	0.1900	0.1864	0.0857	0.2750	0.1227

Table 1 provides descriptive statistics for the variables used in our multivariate analysis for the bid-ask spread models. These descriptive statistics are computed using a sample of securitizing banks. See the Appendix for variable definitions.

information asymmetry among market participants is high, informed traders can exploit their informational advantage at the expense of uninformed traders. The market-makers realize they face an adverse-selection problem and increase the bid-ask spread to protect against expected losses from trading with more informed investors. This argument suggests a positive association between the degree of information asymmetry and bid-ask spreads.

In addition to bid-ask spreads, we also use analyst forecast dispersion as a measure of information uncertainty. Our choice of analyst-related information uncertainty measures is consistent with prior literature (e.g., Leuz 2003; Krishnaswami and Subramaniam 1999) that interprets disagreement among analysts as an indication of the lack of available information about the firm and uses the dispersion of analysts' earnings forecasts as measures of firm-level information asymmetry and information quality. In the context of our study, analyst forecast dispersion is a relevant information uncertainty proxy because analysts focus on firm profitability and securitization directly affects profitability. For the banks in our sample, securitization is an important source of

(Flannery et al. 2004; Van Ness et al. 2001). Prior literature (Van Ness et al. 2001) examining the economic validity of the various models used to decompose bid-ask spreads finds that different models lead to widely varying magnitudes of the estimated adverse-selection component. Van Ness et al. (2001, 96) also find that the adverse-selection estimate is poorly correlated with alternate indicators of asymmetric information and conclude that some of the decomposition models represent merely "noisy transformations of the spread." Therefore, in this study, we chose to use the actual (not decomposed) bid-ask spreads and to include in our multivariate analyses control variables (e.g., the stock exchange listing and a firm's idiosyncratic risk) that proxy for order-processing and inventory-holding costs.

earnings.[12] In addition, financial analysts consistently rate securitization-related disclosure as high in importance, yet assign low ratings for securitization disclosure quality (CFA Institute 2008).[13]

Recourse Uncertainty Measures

Ideally, we would use a direct measure of uncertainty that various market participants face with respect to the true extent of recourse obligations. However, such a measure is not directly observable. Therefore, in our empirical analysis we select several measures that we expect to be correlated with the uncertainty about the extent of recourse. These measures are intended to capture uncertainly about the need for future support for troubled securitization transactions and about banks' willingness to provide such support.

First, we select a bank's securitization volume (*ABS*) and securitization income (*SECINC*) because market participants likely find it more difficult to assess the anticipated need for future support for banks with a larger securitization volume/income. In the presence of higher securitization levels, market participants face greater information uncertainty with respect to the availability and the extent of suitable reinvestment opportunities for the larger amounts of securitization-generated cash. Typically, banks recycle securitization proceeds by using the cash to generate or purchase the same types of assets that were originally securitized (e.g., mortgage loans) and then securitize the new assets as well. Prior studies (Dell'Ariccia et al. 2008; Keys et al. 2009) find that, in an attempt to generate enough loans to maintain a high securitization volume and securitization fees, banks lower their lending standards. In this context, opportunities to reinvest securitization proceeds and lending standards are important in assessing the quality of securitized assets and thus, the need for future support. However, bank outsiders face increased information uncertainty in the presence of larger securitization volume/income because investment opportunities and lending standards are not directly observable to outsiders. In addition, banks with higher securitization volume/income are typically more likely to engage in more complex and more heterogeneous securitization transactions and have a wider range of possible recourse obligations, leaving information users exposed to greater information uncertainty.

Second, we use two proxies for the quality of the securitized assets to capture recourse uncertainty: (1) the proportion of nonperforming loans associated with securitized loans (*NPL_SEC*) and (2) the charge-offs associated with securitized loans (*CHOFF_SEC*). Market participants likely find it more difficult to assess banks' willingness and methods to offer future support to their troubled securitization transactions in the presence of a higher anticipated need for support. In deciding whether to offer voluntary support, banks trade off the reputation benefits for the costs of such support (Vermilyea et al. 2008). When the expected levels of necessary future support are higher, banks get closer to the point where the costs outweigh the benefits. Because outside information users cannot observe the break-even point, they are likely to find it more difficult to assess banks' willingness to offer implicit recourse when the expectations about necessary future support are high. In addition, when the need for the future support is high, banks could be forced to take more types of recourse measures to support their securitization

---

[12] Securitizing banks' reported earnings depend on the volume and profitability of their securitizations because, at the date a securitization transaction takes place, sale accounting requires banks to derecognize the securitized assets, record the cash proceeds, and record a gain or a loss from securitization. Transaction and servicing fees generated through securitization are also important components of total earnings. In our sample, the proportion of income coming from securitization activities (i.e., securitization gains and losses and servicing fees scaled by total net income) represents, on average, about 22 percent of total bank income (median value about 4 percent).

[13] The importance of securitization activities from analysts' point of view is also supported by anecdotal evidence. For the banks in our sample, we find anecdotal evidence in conference call transcripts that, in some cases, analysts ask clarifying questions about the magnitude of the expected securitization gains mentioned by bank managers, or managers provide additional details about the magnitude of expected securitization recourse.

transactions.[14] Consequently, information users are likely to be exposed to greater information uncertainty in the presence of a greater need for support, if a greater need for recourse is associated with a heterogeneous mix of recourse actions.

Also, we use the volume of asset-backed securities retained by securitizing banks on their balance sheets (i.e., retained interests, *RET_INT*) to capture uncertainty about explicit recourse. For securitizing banks, information users can only observe the reported levels of the retained interests, but they do not have detailed information about the retained interests' level of subordination, credit enhancement features, and other details of various heterogeneous securitization transactions necessary to infer the expected future support. The argument that information users can only observe aggregate reported levels of securities holds for other types of on-balance sheet securities as well. However, compared to other securities, retained interests have several features that make them subject to greater information uncertainly. The value of the retained interests is closely intertwined with the value of the pool of securitized assets transferred to the SPEs, representing a highly concentrated risk position. That is, when the pool of transferred assets suffers losses, the retained interests bear significantly more than their proportional share of the losses because of their first-loss position. In addition, retained interests are very illiquid because there is no active market for these securities. Therefore, their valuation depends heavily on managerial assumptions, leaving outside investors exposed to high levels of information asymmetry (Dechow and Shakespeare 2009; Dechow et al. 2010). High levels of retained interests also imply a higher upper limit of possible explicit recourse, which, in turn, likely leads to more uncertainly about the extent of risk transfer achieved through securitization.

It is possible that each of the five recourse uncertainty proxies captures only specific aspects of the underlying concept of interest (i.e., the uncertainty about the extent of risk transfer) and, therefore, includes measurement errors. To mitigate this concern, we use a principal component analysis to extract the relevant information in these measures and compute a composite recourse uncertainty factor.

The Impact of the Difficulty in Estimating Recourse on Bank Information Uncertainty

To test our hypothesis that the difficulty in estimating recourse obligations is positively associated with bank information uncertainty, we regress the two information uncertainty proxies (bid-ask spreads and analyst forecast dispersion) on our recourse uncertainty factor and control variables. Our regressions are specified as follows (firm and quarter subscripts omitted):

$$\begin{aligned} UNCERTAINTY = {} & a_0 + a_1 RECOURSE + a_2 NETLNS + a_3 PROFIT + a_4 MVLEV + a_5 SIZE \\ & + a_6 STDRET + (a_7 PRICE + a_8 NASDAQ + a_9 TURNOVER) + a_{10} FOLLOW \\ & + a_{11} DERIVATIVE + a_{12} CHGOFF\_ONBS + a_{13} NPL\_ONBS + a_{14} GAP + e. \end{aligned} \tag{1}$$

UNCERTAINTY is defined, in turn, as either the bid-ask spread (*SPREAD*) or analyst forecast dispersion (*DISPERSION*). Our main variable of interest in Equation (1) is *RECOURSE*. This variable is the first factor from a principal component analysis using *ABS*, *SECINC*, *NPL_SEC*, *CHOFF_SEC*, and *RET_INT*. If our hypothesis holds, then we expect a positive coefficient on the recourse factor.

[14] These various types of actions can have different future implications for the securitizing bank. For example, to support its securitization structures, a bank can sell assets to a securitization trust (or other SPE) at a discount from the price specified in the securitization documents, which is typically par value. At the same time, the bank can fund credit enhancements beyond contractual requirements. Arguably, the costs of the former type of implicit recourse measure are easier to assess than the costs of the latter type.

In Equation (1), the variables included in parentheses are relevant only for the model in which the bid-ask spread is the dependent variable. Prior research finds that firm size and analyst following are associated with information uncertainty. Therefore, we include these control variables in our empirical models. We expect firm size and analyst following (*SIZE* and *FOLLOW*) to be negatively associated with our measures of information uncertainty because the greater information availability for larger firms and firms with larger analyst following can reduce information uncertainty. Following prior literature, we use bank stock price, share turnover, and an indicator variable for a firm's stock exchange listing in our spread regression. The indicator variable for the exchange listing (*NASDAQ*) is expected to be positively associated with spreads, since prior research finds that firms listed on the NASDAQ stock exchange have larger bid-ask spreads as compared to the NYSE and AMEX firms. Price is expected to have a negative coefficient in the spread regressions. We also include share turnover (*TURNOVER*) as a control variable in the bid-ask spread model and predict a negative coefficient on turnover, because prior literature finds that spreads decrease with trading volume.

To control for differences in bank asset composition, we include the proportion of on-balance sheet net loans to total assets (*NETLNS*), as loans are by far the largest asset on our sample firms' balance sheet. Because prior studies (e.g., Flannery et al. 2004) find mixed results with respect to whether more loans increase risk, we do not have directional predictions for this variable. In addition, we include the ratio of net income to total assets (*PROFIT*) to capture the effect of banks' profitability on information uncertainty.

One caveat of using proxies associated with the difficulty in estimating recourse is that our recourse uncertainty factor might only capture the securitization recourse level. Because the recourse level is a determinant of a bank's intrinsic risk, our recourse uncertainty measure might capture the effect of securitization on bank intrinsic risk and not the effect on bank information uncertainty. Therefore, in Equation (1), we use controls frequently used in prior literature as proxies for firm risk, such as the standard deviation of returns (*STDRET*) and firm leverage (*MVLEV*).[15] In addition, we include several proxies for bank risk that are specific to the banking industry (*DERIVATIVE*, *CHGOFF_ONBS*, *NPL_ONBS*, and *GAP*). We control for banks' derivative use to account for the possibility that banks make the decision of how much recourse to offer in association with their securitized assets in the broader context of an overall risk-management strategy. Prior research (e.g., Minton et al. 2009) finds that banks are more likely to buy derivative protection if they are engaged in asset securitization. Given the level of complexity and illiquidity of some derivative transactions, it is possible that the use of derivatives also leads to increased bank information uncertainty and thus, we expect a positive coefficient on *DERIVATIVE*. The next two measures (*CHGOFF_ONBS* and *NPL_ONBS*) are intended to capture the riskiness of on-balance sheet loans. If market participants face greater uncertainty in assessing banks that have exposure to riskier on-balance sheet loans, then we expect a positive association between the *CHGOFF_ONBS* and *NPL_ONBS* variables and our measures of information uncertainty. The *GAP* variable captures the interest rate risk stemming from the possible mismatching between assets and liabilities, which will be repriced in the short-run. We expect positive coefficients on all of these risk variables. Note that, to some extent, the bank size and the on-balance sheet net loans also control for differences in risk profile among securitizing banks.

Panel A of Table 2 presents the distributional characteristics of the five measures of recourse uncertainty and the first common factor based on these measures for our sample of securitizing

---

[15] Although, prior information asymmetry literature does not traditionally include beta (i.e., a systematic risk measure) as a control variable in bid-ask spreads and forecast dispersion regressions, in untabulated analysis we add firm beta as an additional risk measure. Our results are robust to the inclusion of this control variable.

TABLE 2

Descriptive Statistics for Recourse Uncertainty Variables (n = 871)

Panel A: Distributional Characteristics of the Recourse Uncertainty Variables

Variables	Mean	Median	Q1	Q3	Std. Dev.
ABS	0.1162	0.0141	0.0000	0.0817	0.2573
NPL_SEC	0.0045	0.0001	0.0000	0.0025	0.0118
CHGOFF_SEC	0.0004	0.0000	0.0000	0.0001	0.0010
SECINC	0.2196	0.0368	0.0100	0.1830	0.4889
RET_INT	0.0010	0.0000	0.0000	0.0007	0.0024
RECOURSE	0.0000	−0.4340	−0.5295	0.1211	1.0000

Panel B: Correlations of the Recourse Uncertainty Variables

Variables	*ABS*	*NPL_SEC*	*CHGOFF_SEC*	*SECINC*	*RET_INT*	*RECOURSE*
ABS	1.00	0.95***	0.42***	0.59***	0.52***	0.92***
NPL_SEC	0.91***	1.00	0.47***	0.59***	0.62***	0.92***
CHGOFF_SEC	0.64***	0.73***	1.00	0.70***	0.73***	0.65***
SECINC	0.54***	0.62***	0.51***	1.00	0.63***	0.75***
RET_INT	0.64***	0.70***	0.82***	0.49***	1.00	0.66***
RECOURSE	0.83***	0.87***	0.70***	0.83***	0.67***	1.00

Panel C: Weights (Scoring Coefficients) on the Five Recourse Uncertainty Measures in the First Common Factor

	ABS	*NPL_SEC*	*CHGOFF_SEC*	*SECINC*	*RET_INT*
Weight	0.2174	0.3956	0.2233	0.1450	0.1650

*** Denotes statistical significance at 1 percent based on two-tailed tests.

Table 2 provides descriptive statistics for the recourse uncertainty variables. These descriptive statistics are computed using a sample of securitizing banks.

Panel A reports the descriptive statistics.

Panel B reports correlations between various recourse uncertainty variables. Pearson (Spearman) correlations are presented above (below) the diagonal.

Panel C presents the weight (standardized scoring coefficients) on each of the five measures in a linear combination of measures and the weights that yield the first-factor scores from our principal component analysis.

See the Appendix for variable definitions.

banks. On average, total securitized assets, nonperforming securitized loans, total charge-offs associated with securitized loans, and total retained interests account for 11.62 percent, 0.45 percent, 0.04 percent, and 0.10 percent of total assets, respectively. Securitization income (including service fees) accounts for 21.96 percent of net income of these banks. Panel B of Table 2 presents Pearson and Spearman correlations for the recourse variables. The correlations between these proxies for implicit recourse and retained interests are high, ranging from 0.42 (Pearson correlation between *ABS* and *CHCOFF_SEC*) to 0.95 (Pearson correlation between *ABS* and *NPL_SEC*). These high correlations are reasonable, given our argument that all five measures represent the same underlying concept, i.e., the extent of recourse related to securitized assets. However, the results in Table 2 show that, although highly correlated, these five variables do not perfectly overlap and likely capture different aspects of the uncertainty of risk transfer. Panel C of Table 2 presents the weights (i.e., the standardized scoring coefficients) on the five recourse uncertainty measures used to compute the first common factor from our principal component analysis. These weights indicate that all five measures contribute substantially to the computation of our recourse uncertainty factor. We retain only the first common factor in the principal component analysis, since this is the only factor with an eigenvalue greater than 1. This first factor retains about 70 percent of the total variance of the original data.

Table 3 presents the results of estimating Equation (1) using a sample of securitizing banks. Model 1 (Model 2) uses the effective bid-ask spread (analyst forecast dispersion) as our proxy for information uncertainty. The findings in Table 3 support our hypothesis that the level of difficulty in estimating securitization recourse is positively associated with bank-level information uncertainty. Consistent with our expectations, the recourse uncertainty factor has statistically significant, positive coefficients in both the spread (t-statistic = 2.10; one-tailed p-value = 0.05) and the analyst forecast dispersion regressions (t-statistic = 1.75; one-tailed p-value = 0.05). All the control variables are either consistent with our predictions or insignificant, with one exception. While we predict a positive coefficient on the charge-offs on balance sheet loans variable (*CHGOFF_ONBS*), we find a negative and marginally significant (t-statistic = −1.74; two-tailed p-value = 0.10) coefficient in the bid-ask spread model.

The Impact of the Information Environment Quality on the Relation between Recourse and Information Uncertainty

Given our finding that, on average, the difficulty in assessing the extent of securitization recourse for securitizing banks is positively associated with bank information uncertainty, we next investigate whether this effect is mitigated for banks that operate in a higher quality information environment. To test this hypothesis, we run the following regression (firm and quarter subscripts omitted):

$$\begin{aligned} UNCERTAINTY = {} & a_0 + a_1 RECOURSE + a_2 RECOURSE * QUALITY\_INFO\_ENV \\ & + a_3 QUALITY\_INFO\_ENV + a_4 NETLNS + a_5 PROFIT + a_6 MVLEV \\ & + a_7 STDRET\ (+ a_8 PRICE + a_9 NASDAQ + a_{10} TURNOVER) \\ & + a_{11} DERIVATIVE + a_{12} CHGOFF\_ONBS + a_{13} NPL\_ONBS + a_{14} GAP + e. \end{aligned} \tag{2}$$

The *QUALITY_INFO_ENV* variable is defined, in turn, as either analyst following (*FOLLOW*), or the level of institutional investor holdings (*INST*), or size (*SIZE*). If our hypothesis holds, then we expect the coefficient on the interaction term between the recourse uncertainty

TABLE 3

The Impact of Recourse Uncertainty on Banks' Bid-Ask Spreads and Analyst Forecast Dispersion

	Predicted Sign	Bid-Ask Spread		Forecast Dispersion	
		Coefficient	**t-statistic**	**Coefficient**	**t-statistic**
Intercept	?	0.0255	(5.19)***	0.0159	(0.75)
RECOURSE	+	0.0013	(2.10)**	0.0069	(1.75)**
NETLNS	?	−0.0044	(−1.27)	0.0011	(0.09)
PROFIT	?	0.0294	(0.40)	0.0029	(0.01)
MVLEV	?	0.0001	(0.85)	0.0018	(2.95)***
SIZE	−	−0.0017	(−3.31)***	−0.0007	(−0.43)
STDRET	+	0.1656	(1.93)**	0.4045	(0.95)
PRICE	−	−0.0001	(−1.26)	NA	NA
NASDAQ	+	0.0016	(1.43)*	NA	NA
TURNOVER	−	−0.0124	(−2.13)**	NA	NA
FOLLOW	−	−0.0019	(−2.07)**	−0.0056	(−2.19)**
DERIVATIVE	+	0.0003	(2.65)***	0.0006	(3.10)***
CHGOFF_ONBS	+	−0.2007	(−1.74)*	−0.5754	(−0.73)
NPL_ONBS	+	0.0900	(1.12)	0.0358	(0.15)
GAP	+	0.0046	(1.80)**	0.0035	(0.25)
Adj-R^2		55.09%		23.27%	
n		871		687	

*, **, *** Denote statistical significance at 10 percent, 5 percent, and 1 percent, respectively, based on one-tailed tests if the coefficient has a predicted sign, and two-tailed tests otherwise.

Table 3 presents the results from the regressions of banks' bid-ask spreads and analyst forecast dispersion on the recourse uncertainty factor and control variables using a sample of securitizing banks. Standard errors are cluster-adjusted by firm and quarter following Petersen (2009) and Gow et al. (2010).

See the Appendix for variable definitions.

factor and the quality of the information environment to be negative (i.e., $a_2 < 0$) and the coefficient on the recourse uncertainty factor to be positive (i.e., $a_1 > 0$).

We select analyst following as a proxy for the quality of the information environment because prior research finds that when analyst following is large, greater competition among analysts leads to more timely price discovery (e.g., Brennan et al. 1993; Brennan and Subrahmanyam 1996) and lower information asymmetries in markets (e.g., Hong et al. 2000; Barth and Hutton 2004; Ellul and Panayides 2008). A separate stream of literature (Lang and Lundholm 1996) finds that analysts are more likely to cover firms with higher quality disclosure policies. We use the level of institutional investor holdings as a second proxy for the quality of the information environment because institutional investors have a higher level of expertise in processing relevant information. Such expertise is particularly important in our setting because the underlying securitization transactions are very complex and thus, by necessity, the securitization accounting and disclosures are technically complex. In addition, we choose bank size as our third measure of information environment quality because larger banks are likely to attract more press coverage and information users'

scrutiny, leading to a more timely and comprehensive information discovery process.[16,17]

The findings presented in Table 4 (Panels A, B, and C) support our hypothesis. The coefficients on the interaction term between the recourse uncertainty factor and the information environment quality are negative and statistically significant for all three measures of information environment quality, for both the bid-ask spread and the forecast dispersion regressions. In addition, supporting our hypothesis, the coefficients on the recourse uncertainty factor are positive and statistically significant. All control variables are either consistent with our predictions or statistically insignificant at conventional significance levels.

Our findings show that, for securitizing banks, there is a positive association between the difficulty in estimating recourse and bank information uncertainty, even after including in our regressions control measures commonly used in the literature as proxies for intrinsic risk. However, to the extent that these typical risk controls measure bank intrinsic risk with error, it could be that our recourse uncertainty factor captures the impact of securitization on bank intrinsic risk, while we spuriously attribute the effect to information uncertainty. The findings presented in Table 4 help alleviate this concern. On the one hand, if our recourse uncertainty factor captures the information effect of securitization, then we expect the impact of recourse measures on bank information uncertainty to be mitigated for banks that operate in a high-quality information environment. On the other hand, if the factor captures only the impact of recourse on bank risk, then we would not expect the relation between recourse and information uncertainty to vary systematically with the quality of the information environment. The results in Table 4 offer support for the former expectation.

Following Fernandes and Ferreira (2008), to further rule out the alternative explanation that our recourse uncertainty factor captures only the impact of securitization on bank intrinsic risk and not the impact on information uncertainty, we also investigate whether our quality-of-information-environment results are robust to controlling for the impact of intrinsic risk on the relation between our recourse uncertainty measure and bank information uncertainty. Specifically we estimate the following equation (firm and quarter subscripts omitted):

$$\begin{aligned} \mathit{UNCERTAINTY} = {} & a_0 + a_1\mathit{RECOURSE} + a_2\mathit{RECOURSE} * \mathit{QUALITY\_INFO\_ENV} \\ & + a_3\mathit{RECOURSE} * \mathit{STDRET} + a_4\mathit{QUALITY\_INFO\_ENV} + a_5\mathit{NETLNS} \\ & + a_6\mathit{PROFIT} + a_7\mathit{MVLEV} + a_8\mathit{STDRET}\ (+ a_9\mathit{PRICE} + a_{10}\mathit{NASDAQ} \\ & + a_{11}\mathit{TURNOVER}) + a_{12}\mathit{DERIVATIVE} + a_{13}\mathit{CHGOFF\_ONBS} \\ & + a_{14}\mathit{NPL\_ONBS} + a_{15}\mathit{GAP} + e. \end{aligned} \tag{3}$$

[16] Note that there are two types of information that affect a bank's information environment and can help mitigate the uncertainty about the extent of recourse: (1) securitization transaction-specific information about the quality of securitized assets and the implicit and explicit guaranties of support offered by securitizing banks; and (2) general bank information about investment opportunities, lending standards, alternative sources of financing, and bank reputation concerns, etc. Ideally, we would assess the quality of a bank information environment by investigating both these types of information. In untabulated analysis, we hand-collect securitization transaction disclosures for our sample banks from the footnotes to the financial statements included in the 10-K filings. Consistent with complaints by investors and financial analysts, we find that the securitization-related footnote disclosure is relatively standardized, making it very difficult to capture cross-sectional variation in the quality of securitization footnote disclosure for our sample firms. Therefore, in our empirical analyses we focus on bank-level information quality aspects.

[17] In a sensitivity test, we use principal component analysis to develop an information environment quality factor based on size, institutional holding, and analyst following. We use this information quality factor to replace each individual information environment measure and rerun the analyses in Tables 4 and 5. We find that our results are robust to this alternative specification.

TABLE 4

The Impact of Information Environment Quality on the Relation between Recourse Uncertainty and Banks' Bid-Ask Spreads and Analyst Forecast Dispersion

Panel A: Using Analyst Following (*FOLLOW*) as a Measure of the Quality of the Information Environment

		Bid-Ask Spread		Forecast Dispersion	
	Predicted Sign	Coefficient	t-statistic	Coefficient	t-statistic
Intercept	?	0.0267	(5.18)***	0.0159	(0.70)
RECOURSE	+	0.0033	(2.94)***	0.0340	(3.52)***
*RECOURSE * FOLLOW*	−	−0.0010	(−1.82)**	−0.0104	(−4.02)***
NETLNS	?	−0.0052	(−1.44)	−0.0021	(−0.16)
PROFIT	?	−0.0716	(−0.90)	−0.5845	(−1.84)*
MVLEV	?	0.0001	(0.64)	0.0016	(2.37)**
SIZE	−	−0.0016	(−3.29)***	0.0001	(0.03)
STDRET	+	0.1463	(1.73)**	0.3036	(0.80)
PRICE	−	−0.0001	(−1.31)	NA	NA
NASDAQ	+	0.0018	(1.57)*	NA	NA
TURNOVER	−	−0.0160	(−2.92)***	NA	NA
FOLLOW	−	−0.0021	(−2.35)***	−0.0055	(−1.91)**
DERIVATIVE	+	0.0002	(2.10)**	0.0003	(1.09)
CHGOFF_ONBS	+	−0.0794	(−0.58)	0.1491	(0.28)
NPL_ONBS	+	0.1053	(1.28)	0.0921	(0.46)
GAP	+	0.0042	(1.71)*	−0.0014	(−0.11)
Adj-R^2		56.70%		28.35%	
n		871		687	

Panel B: Using Institutional Holding (*INST*) as a Measure of the Quality of the Information Environment

		Bid-Ask Spread		Forecast Dispersion	
	Predicted Sign	Coefficient	t-statistic	Coefficient	t-statistic
Intercept	?	0.0250	(5.06)***	0.0095	(0.34)
RECOURSE	+	0.0032	(3.76)***	0.0210	(2.16)**

(continued on next page)

Panel B: Using Institutional Holding (*INST*) as a Measure of the Quality of the Information Environment

	Predicted Sign	Bid-Ask Spread Coefficient	Bid-Ask Spread t-statistic	Forecast Dispersion Coefficient	Forecast Dispersion t-statistic
*RECOURSE * INST*	−	−0.0039	(−2.37)***	−0.0262	(−1.97)**
INST	?	−0.0034	(−1.08)	0.0210	(1.13)
NETLNS	?	−0.0038	(−1.06)	0.0077	(0.47)
PROFIT	?	−0.0385	(−0.65)	−0.2840	(−0.77)
MVLEV	?	0.0000	(−0.14)	0.0020	(2.41)**
SIZE	−	−0.0013	(−2.83)***	−0.0007	(−0.36)
STDRET	+	0.1714	(2.35)***	0.3549	(0.90)
PRICE	−	−0.0001	(−1.35)*	NA	NA
NASDAQ	+	0.0013	(1.25)	NA	NA
TURNOVER	−	−0.0132	(−2.38)***	NA	NA
FOLLOW	−	−0.0015	(−1.39)*	−0.0094	(−2.76)***
DERIVATIVE	+	0.0002	(2.31)**	0.0005	(1.83)**
CHGOFF_ONBS	+	−0.1352	(−1.44)	−0.1127	(−0.18)
NPL_ONBS	+	0.0255	(0.47)	0.2079	(0.97)
GAP	+	0.0019	(0.81)	0.0043	(0.34)
Adj-R^2		59.01%		27.11%	
n		871		687	

Panel C: Using Size (*SIZE*) as a Measure of the Quality of the Information Environment

	Predicted Sign	Bid-Ask Spread Coefficient	Bid-Ask Spread t-statistic	Forecast Dispersion Coefficient	Forecast Dispersion t-statistic
Intercept	?	0.0262	(5.07)***	0.0159	(0.76)
RECOURSE	+	0.0049	(1.89)**	0.0453	(2.14)**
*RECOURSE * SIZE*	−	−0.0004	(−1.40)*	−0.0038	(−1.94)**
NETLNS	?	−0.0045	(−1.31)	0.0024	(0.18)
PROFIT	?	−0.0256	(−0.36)	−0.2637	(−0.74)
MVLEV	?	0.0001	(0.84)	0.0018	(2.81)***
SIZE	−	−0.0017	(−3.21)***	−0.0010	(−0.58)
STDRET	+	0.1499	(1.74)**	0.3372	(0.86)

(continued on next page)

Panel C: Using Size (*SIZE*) as a Measure of the Quality of the Information Environment

	Predicted Sign	Bid-Ask Spread Coefficient	Bid-Ask Spread t-statistic	Forecast Dispersion Coefficient	Forecast Dispersion t-statistic
PRICE	–	<−0.0001	(−1.35)*	NA	NA
NASDAQ	+	0.0017	(1.54)*	NA	NA
TURNOVER	–	−0.0150	(−2.87)***	NA	NA
FOLLOW	–	−0.0019	(−2.05)**	−0.0041	(−1.37)*
DERIVATIVE	+	0.0003	(2.62)***	0.0006	(2.62)***
CHGOFF_ONBS	+	−0.1238	(−0.93)	−0.2052	(−0.30)
NPL_ONBS	+	0.0931	(1.17)	0.0294	(0.14)
GAP	+	0.0044	(1.77)**	0.0006	(0.04)
Adj-R^2		55.89%		24.56%	
n		871		687	

*, **, *** Denote statistical significance at 10 percent, 5 percent, and 1 percent, respectively, based on one-tailed tests if the coefficient has a predicted sign, and two-tailed tests otherwise.

Table 4 presents the results from the regressions of banks' bid-ask spreads and analyst forecast dispersion on the recourse uncertainty factor, the recourse uncertainty factor interacted with information environment quality proxies, and control variables. The analyses in Panels A, B, and C use the number of analysts following, the percentage of institutional holding, and the firm size, respectively, to measure the quality of the information environment. Standard errors are cluster-adjusted by firm and quarter following Petersen (2009) and Gow et al. (2010)

See the Appendix for variable definitions.

All variables are defined as before. Compared with Equation (2), Equation (3) includes an additional interaction term between recourse uncertainty and intrinsic risk.

If the factor captures the information effect of securitization, then we expect the impact of recourse measures on bank information uncertainty to be mitigated for banks that operate in a higher quality information environment even after adding an interaction term between recourse and intrinsic risk (as measured by the standard deviation of returns). In other words, we expect the coefficient on the interaction term between the recourse uncertainty factor and the quality of the information environment to be negative (i.e., $a_2 < 0$) and the coefficient on the recourse uncertainty factor to be positive (i.e., $a_1 > 0$) even after controlling for *RECOURSE* $*$ *STDRET*.

As shown in Table 5 (Panels A, B, and C), consistent with our prediction, the coefficients on the interaction term between the recourse uncertainty factor and the information environment quality are negative and statistically significant for all three measures of information environment quality.[18] In addition, the coefficients on the recourse uncertainty factor remain positive and statistically significant in all of these alternative specifications. However, the coefficients on the interaction term *RECOURSE* $*$ *STDRET* are not statistically significant at conventional levels for any of our specifications, indicating that the relation between the factor and bank information uncertainty does not systematically vary with the level of bank intrinsic risk (as measured by the standard deviation of stock returns). These findings further alleviate the concern that our findings capture only the impact of securitization recourse on bank risk.

The Impact of Securitization on Information Uncertainty

So far, our findings indicate that the difficulty in estimating recourse has a significant impact on bank information uncertainty for a sample of securitizing banks and this effect is mitigated for banks that operate in a higher quality information environment. An interesting, related question is whether securitization leaves banks that engage in this type of financing transactions more or less exposed to information uncertainty as compared to banks that do not securitize assets. Since securitization transactions can impact information uncertainty through several mechanisms, it is not clear *ex ante* whether securitizing banks face higher or lower levels of information uncertainty compared to banks that do not securitize assets. First, securitizing banks are exposed to uncertainty about whether the risks associated with the securitized assets have been transferred from the originating bank to investors in asset-backed securities. Second, compared to non-securitizing banks, banks that undertake securitization transactions are exposed to uncertainty regarding the effect of reinvesting the considerable amounts of securitization proceeds. Market participants face uncertainty about the availability of suitable reinvestment opportunities for the cash proceeds. These two mechanisms would suggest higher information uncertainty for securitizing compared to non-securitizing banks. Third, in the securitization process, banks have to satisfy additional disclosure requirements for the transferred assets.[19] By providing the additional required disclosure, securitization can lead to increased transparency with respect to not only the securitized assets, but also the securitizing firms in general (Foley et al. 1999; Berger and Udell 1995; Cantor and Rouyer 2000). Fourth, it is possible that the difficulty in estimating recourse is so great that

---

[18] When we control for bank intrinsic risk using beta as opposed to the standard deviation of returns, we obtain similar results.

[19] For example, a prospectus issued by an SPE to register asset-backed securities with the SEC can include information about initial principal balances for various classes of securities issued, pass-through rates, principal type, interest type, initial credit ratings, etc. It can also include a discussion of various risk factors and information about servicers and servicing fees, such as a description of the different components of the servicing fees, various types of servicer expenses, supplemental servicing fee schedules, etc. Because in many securitizations, the securitizing bank also retains the role of servicer, the prospectus information about the servicer is directly relevant to bank outsiders in assessing the securitizing bank's business model and sources of revenue.

TABLE 5

The Impact of Information Environment Quality and Risk on the Relation between Recourse Uncertainty and Banks' Bid-Ask Spreads and Analyst Forecast Dispersion

Panel A: Using Analyst Following (*FOLLOW*) as a Measure of the Quality of the Information Environment

	Predicted Sign	Bid-Ask Spread		Forecast Dispersion	
		Coefficient	t-statistic	Coefficient	t-statistic
Intercept	?	0.0268	(5.15)***	0.0157	(0.70)
RECOURSE	+	0.0038	(2.30)**	0.0328	(3.65)***
*RECOURSE * FOLLOW*	–	−0.0010	(−1.81)**	−0.0104	(−4.07)***
*RECOURSE * STDRET*	?	−0.0167	(−0.53)	0.0520	(0.68)
NETLNS	?	−0.0052	(−1.43)	−0.0020	(−0.15)
PROFIT	?	−0.0727	(−0.91)	−0.5854	(−1.84)*
MVLEV	?	0.0001	(0.54)	0.0016	(2.38)**
SIZE	–	−0.0016	(−3.31)***	0.0001	(0.04)
STDRET	+	0.1511	(1.74)**	0.2911	(0.78)
PRICE	–	−0.0001	(−1.29)	NA	NA
NASDAQ	+	0.0018	(1.58)*	NA	NA
TURNOVER	–	−0.0158	(−3.01)***	NA	NA
FOLLOW	–	−0.0021	(−2.35)**	−0.0055	(−1.92)**
DERIVATIVE	+	0.0002	(2.09)**	0.0003	(1.11)
CHGOFF_ONBS	+	−0.0817	(−0.60)	0.1559	(0.29)
NPL_ONBS	+	0.1060	(1.28)	0.0906	(0.45)
GAP	+	0.0041	(1.69)**	−0.0015	(−0.11)
Adj-R^2		56.68%		28.28%	
n		871		687	

Panel B: Using Institutional Holding (*INST*) as a Measure of the Quality of the Information Environment

	Predicted Sign	Bid-Ask Spread		Forecast Dispersion	
		Coefficient	t-statistic	Coefficient	t-statistic
Intercept	?	0.0251	(5.08)***	0.0091	(0.33)

(continued on next page)

Panel B: Using Institutional Holding (*INST*) as a Measure of the Quality of the Information Environment

	Predicted Sign	Bid-Ask Spread Coefficient	Bid-Ask Spread t-statistic	Forecast Dispersion Coefficient	Forecast Dispersion t-statistic
RECOURSE	+	0.0034	(4.77)***	0.0194	(2.29)**
*RECOURSE * INST*	–	−0.0040	(−2.57)***	−0.0262	(−1.97)**
*RECOURSE * STDRET*	?	−0.0069	(−0.38)	0.0759	(0.71)
INST	?	−0.0034	(−1.11)	0.0213	(1.14)
NETLNS	?	−0.0038	(−1.06)	0.0079	(0.49)
PROFIT	?	−0.0385	(−0.65)	−0.2866	(−0.77)
MVLEV	?	0.0000	(−0.19)	0.0020	(2.42)**
SIZE	–	−0.0013	(−2.83)***	−0.0006	(−0.35)
STDRET	+	0.1736	(2.24)**	0.3363	(0.87)
PRICE	–	−0.0001	(−1.35)*	NA	NA
NASDAQ	+	0.0013	(1.25)	NA	NA
TURNOVER	–	−0.0131	(−2.47)**	NA	NA
FOLLOW	–	−0.0015	(−1.39)*	−0.0095	(−2.76)***
DERIVATIVE	+	0.0002	(2.31)**	0.0005	(1.85)**
CHGOFF_ONBS	+	−0.1370	(−1.41)	−0.1002	(−0.16)
NPL_ONBS	+	0.0257	(0.47)	0.2070	(0.97)
GAP	+	0.0018	(0.80)	0.0043	(0.33)
Adj-R^2		59.29%		27.10%	
n		871		687	

Panel C: Using Size (*SIZE*) as a Measure of the Quality of the Information Environment

	Predicted Sign	Bid-Ask Spread Coefficient	Bid-Ask Spread t-statistic	Forecast Dispersion Coefficient	Forecast Dispersion t-statistic
Intercept	?	0.0263	(5.03)***	0.0159	(0.75)
RECOURSE	+	0.0055	(1.81)**	0.0450	(2.19)**
*RECOURSE * SIZE*	–	−0.0004	(−1.42)*	−0.0038	(−1.96)**
*RECOURSE * STDRET*	?	−0.0142	(−0.48)	0.0099	(0.21)
NETLNS	?	−0.0045	(−1.30)	0.0024	(0.18)
PROFIT	?	−0.0265	(−0.37)	−0.2631	(−0.74)

(continued on next page)

Panel C: Using Size (*SIZE*) as a Measure of the Quality of the Information Environment

	Predicted Sign	Bid-Ask Spread Coefficient	Bid-Ask Spread t-statistic	Forecast Dispersion Coefficient	Forecast Dispersion t-statistic
MVLEV	?	0.0001	(0.77)	0.0018	(2.82)***
SIZE	–	−0.0017	(−3.21)***	−0.0010	(−0.57)
STDRET	+	0.1536	(1.75)**	0.3350	(0.85)
PRICE	–	−0.0001	(−1.33)*	NA	NA
NASDAQ	+	0.0018	(1.55)*	NA	NA
TURNOVER	–	−0.0148	(−2.92)***	NA	NA
FOLLOW	–	−0.0019	(−2.05)**	−0.0041	(−1.38)*
DERIVATIVE	+	0.0003	(2.62)***	0.0006	(2.63)***
CHGOFF_ONBS	+	−0.1251	(−0.94)	−0.2050	(−0.30)
NPL_ONBS	+	0.0933	(1.18)	0.0291	(0.14)
GAP	+	0.0043	(1.77)**	0.0006	(0.04)
Adj-R^2		55.84%		24.45%	
n		871		687	

*, **, *** Denote statistical significance at 10 percent, 5 percent, and 1 percent, respectively, based on one-tailed tests if the coefficient has a predicted sign, and two-tailed tests otherwise.

Table 5 presents results from regressions of banks' bid-ask spreads and analyst forecast dispersion on the recourse uncertainty factor, the recourse uncertainty factor interacted with information environment quality proxies and with bank intrinsic risk, as well as control variables. The analyses in Panels A, B, and C use the number of analysts following, the percentage of institutional holdings, and the firm size, respectively, to measure the quality of the information environment. Standard errors are cluster-adjusted by firm and quarter following Petersen (2009) and Gow et al. (2010).

See the Appendix for variable definitions.

information users protect themselves against high uncertainty by assuming minimal risk transfer for all securitized assets (i.e., assuming that all securitizations are, in fact, secured borrowings).[20] In this case, we would expect no difference in information uncertainty between securitizing and non-securitizing banks. Given these opposite expectations, it is unclear *ex ante* whether the net effect of securitization would be an increase, a decrease, or no change in information uncertainty relative to non-securitizing banks.

To assess the effect of securitization on information uncertainty, we match our sample of securitizing banks by size, risk level (i.e., standard deviation of returns), and quarter with a sample of non-securitizing banks. This matching allows us to investigate whether, holding constant important bank characteristics such as size and risk, the propensity to undertake securitization transactions affects bank information uncertainty. We then regress each of the two information uncertainty proxies (the bid-ask spread and the analyst forecast dispersion) on an indicator variable (*ABSD*) that takes a value of 1 (0) for securitizing (non-securitizing) banks and the control variables used in Equation (1). The results from this analysis (untabulated) indicate that the coefficients for *ABSD* in both the bid-ask spread and the forecast dispersion models are positive and statistically significant. These findings suggest that securitizing banks have higher levels of information uncertainty as compared to banks that do not use this financing technique.

It is possible that a bank's level of information uncertainty and its securitization activities are endogenously determined. The information uncertainty faced by banks could also influence their decision of whether to securitize. The literature cites immediate access to capital and the diversification of funding sources as some of the main benefits of securitization. To the extent that banks with high levels of information uncertainty find it more difficult and/or costly to access other sources of capital, they may be more likely to undertake asset securitization transactions to finance their operations.

To control for this possible endogeneity problem, we estimate a two-equation system, wherein we model the decision to securitize assets and the bank information uncertainty simultaneously.[21] Hausman tests reject the null hypothesis of no endogeneity for both our measures of information uncertainty. Our (untabulated) findings based on this system of equations provide support for the hypothesis that securitizing banks have higher information uncertainty (specifically, higher bid-ask spreads and analyst forecast dispersion) compared to non-securitizing banks.[22]

VI. SUMMARY AND CONCLUSION

In this study, we investigate the effect of bank asset securitization on information uncertainty. We first hypothesize that, for securitizing banks, difficulty in estimating securitization recourse can lead to higher information uncertainty among investors with different access to information and different abilities to process information. We then investigate whether this effect can be mitigated in settings where the quality of the information environment is high. Finally, we compare the information uncertainty levels between securitizing and non-securitizing banks, after controlling for other determinants of information uncertainty and for endogeneity.

We find that our empirical proxies for the difficulty in assessing recourse are positively associated with bank information uncertainty. However, this effect is mitigated for securitizing banks that operate in a higher quality information environment (as measured by larger analyst

[20] Prior theoretical studies (e.g., Cao et al. 2005) show that investors will adopt a worst-case scenario view when faced with large levels of uncertainty.

[21] While in our empirical tests of comparing securitizing and non-securitizing banks we attempt to control for bank-relevant characteristics, we cannot completely rule out the possibility that unobserved (and uncontrolled for) differences between securitizing and non-securitizing banks are driving our findings.

[22] The results for both the matched sample analysis and the 2SLS analysis are available from the authors on request.

following, larger institutional ownership, and greater bank size). Our findings also indicate that banks that engage in securitization transactions have higher bid-ask spreads and analyst forecast dispersion compared with banks that do not engage in this type of financing transactions.

This study contributes to the literature by investigating the implications of securitization transactions and securitization recourse on information uncertainty. Our findings contribute to a better understanding of how market participants differ in their assessment of complex transactions such as securitizations. Specifically, our findings imply that inadequate disclosures can lead to higher information uncertainty for securitizing banks, and that higher information quality can help mitigate this uncertainty. These findings provide support for FASB's recent efforts to improve securitization disclosures (FASB 2008; ASC 860-10, formerly SFAS No. 166, FASB 2009a, 2009b).

APPENDIX
VARIABLE DEFINITIONS

SPREAD	=	average of daily effective bid-ask spreads during quarter *t*, where the effective bid-ask spread is measured as two times the absolute value of the difference between the daily ending trading price and the average of the ask and bid scaled by the average of bid and ask (from TAQ consolidated trades and consolidated quotes files);
DISPERSION	=	one-year-ahead analyst forecast dispersion calculated in the middle of the second month during quarter *t*, where dispersion is measured as the standard deviation of forecasts scaled by mean of forecasts (from I/B/E/S summary file);
ABSD	=	an indicator variable that takes the value of 1 if a firm has any securitized assets at the end of quarter *t*, and 0 otherwise;
NETLNS	=	total loans, net of the allowance for loan and lease losses plus customers' liabilities on outstanding acceptances scaled by total assets at end of quarter *t*;
PROFIT	=	ratio of net income to total assets at the end of the quarter *t*;
MVLEV	=	ratio of total liability to book value of equity at the end of quarter *t*;
SIZE	=	log of total assets at the end of quarter *t*;
STDRET	=	return volatility measured as the standard deviation of daily returns during quarter *t*;
PRICE	=	price per share at the end of quarter *t*;
NASDAQ	=	dummy variable that takes the value of 1 if a firm is listed on the NASDAQ, and 0 if the firm is listed on the NYSE or the AMEX;
TURNOVER	=	average daily turnover during quarter *t*, where turnover is measured as trading volume scaled by market capitalization at the end of each day (from the CRSP daily stock file);
FOLLOW	=	log of 1 plus the number of analysts following measured in the middle of the last month during quarter *t* (from the I/B/E/S summary file); if a firm is not covered by an analyst, *FOLLOW* = 0;
DERIVATIVES	=	total notional amount of derivative contracts scaled by total assets at the end of quarter *t*;
CHGOFF_ONBS	=	charge-offs on balance sheet loans scaled by total assets at the end of quarter *t*;
NPL_ONBS	=	past due on balance sheet loans scaled by total assets at the end of quarter *t* (where the past due loans equal loans past due 30 through 89 days plus loans past due 90 days or more);

(continued on next page)

GAP	=	absolute value of the difference between assets and liabilities expected to reprice within the next 12 months scaled by total assets at the end of quarter *t*;
ABS	=	total securitized assets scaled by total assets at the end of quarter *t*;
SECINC	=	securitization income and servicing fees scaled by total net income for quarter *t*;
NPL_SEC	=	total nonperforming securitized loans scaled by total assets at the end of quarter *t*;
CHOFF_SEC	=	total charge-offs for securitized loans scaled by total assets at the end of quarter *t*;
RET_INT	=	total retained interest from all asset securitizations scaled by total assets at the end of quarter *t*; and
RECOURSE	=	first factor from a principal component analysis using *ABS, SECINC, NPL_SEC, CHOFF_SEC*, and *RET_INT*.

REFERENCES

Amihud, Y., and H. Mendelson. 1986. Asset pricing and the bid-ask spread. *Journal of Financial Economics* 17 (2): 223–249.

———, and ———. 1989. The effect of beta, bid-ask spread, residual risk and size on stock returns. *The Journal of Finance* 44 (2): 479–486.

Anderson, E., E. Ghysels, and J. Juergens. 2009. The impact of risk and uncertainty on expected returns. *Journal of Financial Economics* 94 (2): 233–263.

Barth, M., G. Clinch, and T. Shibano. 2003. Market effects of recognition and disclosure. *Journal of Accounting Research* 41 (4): 581–609.

———, and A. P. Hutton. 2004. Analysts earnings forecast revisions and the pricing of accruals. *Review of Accounting Studies* 9 (1): 59–96.

Barry, C., and S. Brown. 1985. Differential information and security market equilibrium. *Journal of Financial and Quantitative Analysis* 20 (4): 407–422.

Berger, A., and G. Udell. 1995. Relationship lending and lines of credit in small firm finance. *The Journal of Business* 68 (3): 358–381.

Brennan, M. J., N. Jegadeesh, and B. Swaminathan. 1993. Investment analysis and the adjustment of stock prices to common information. *Review of Financial Studies* 6 (4): 799–824.

———, and A. Subrahmanyam. 1996. Market microstructure and asset pricing: On the compensation for illiquidity in stock returns. *Journal of Financial Economics* 41 (3): 441–464.

Calomiris, C., and J. Mason. 2004. Credit card securitization and regulatory arbitrage. *Journal of Financial Services Research* 26 (1): 5–27.

Cantor, R., and S. Rouyer. 2000. Another perspective on credit risk transfer and assets securitization. *The Journal of Risk Finance* 1 (2): 37–47.

Cao, H. H., T. Wang, and H. H. Zhang. 2005. Model uncertainty, limited market participation, and asset prices. *Review of Financial Studies* 18 (4): 1219–1251.

CFA Institute. 2008. Comment letter to Proposed FASB Staff Position. No. FAS 140-e and FIN46(R)-e, Disclosure about transfers of financial assets and interests in variable interest entities.

Chen, W., C. Liu, and S. Ryan. 2008. Characteristics of securitization that determine issuers' retention of the risk of the securitized assets. *The Accounting Review* 83 (5): 1181–1215.

Chordia, T., A. Subrahmanyam, and V. Anshuman. 2001. Trading activity and expected stock returns. *Journal of Financial Economics* 59 (1): 3–32.

Clarkson, P., J. Guedes, and R. Thompson. 1996. On the diversification, observability and measurement of estimation risk. *Journal of Financial and Quantitative Analysis* 31 (1): 69–84.

Clarkson, B., K. Becker, and D. Fanger. 2002. *Securitization and its Effect on the Credit Strength of Companies: Moody's Perspective 1987–2002*. New York, NY: Moody's Investors Service.

Dechow, P., and C. Shakespeare. 2009. Do managers time securitization transactions to obtain accounting benefits? *The Accounting Review* 84 (1): 99–132.

———, L. Myers, and C. Shakespeare. 2010. Fair value accounting and gains from asset securitizations: A convenient earnings management tool with compensation side-benefits. *Journal of Accounting and Economics* 49 (1-2): 2–25.

Dell'Ariccia, G., D. Igan, and L. Laeven. 2008. Credit booms and lending standards: Evidence from the subprime mortgage market. Working paper, International Monetary Fund.

Easley, D., and M. O'Hara. 2004. Information and the cost of capital. *The Journal of Finance* 59 (4): 1553–1583.

Ellul, A., and M. Panayides. 2008. Do financial analysts restrain insiders' informational advantage? Working paper, Indiana University Bloomington.

Fernandes, N., and M. A. Ferreira. 2008. Does cross-listing improve the information environment? *Journal of Financial Economics* 88 (2): 216–244.

Financial Accounting Standards Board (FASB). 2000. *Accounting for Transfers and Servicing of Financial Assets and Extinguishments of Liabilities*. Statement of Financial Accounting Standards No. 140. Norwalk, CT: (FASB).

———. 2002. FASB staff reviews compliance by Fortune 1000 companies with securitization accounting standard FAS 140.

———. 2008. Disclosure about transfers of financial assets and interests in variable interest entities. Proposed FASB Staff Position No. FAS 140-e and FIN 46(R)-e. Norwalk, CT: FASB.

———. 2009a. *Accounting for Transfers of Financial Assets—an Amendment of FASB Statement No. 140*. Statement of Financial Accounting Standards No. 166. Norwalk, CT: FASB.

———. 2009b. Accounting Standards Codification 860, available at http://asc.fasb.org/topic&nav_type=topic_page%26trid=2197590. Norwalk, CT: FASB.

Flannery, M., S. Kwan, and M. Nimalendran. 2004. Market evidence on the opaqueness of banking firms' assets. *Journal of Financial Economics* 71 (3): 419–460.

Foley, T., R. Cantor, and M. Anderson. 1999. The evolution of Moody's views on securitization. *Moody's Special Report (May).*

Gorton, G., and N. Souleles. 2006. Special purpose vehicles and securitization. In *The Risks of Financial Institutions*, edited by R. Stulz and M. Carey, 549–602, Chicago, IL: University of Chicago Press.

Gow, I., G. Ormazabal, and D. Taylor. 2010. Correcting for cross-sectional and time-series dependence in accounting research. *The Accounting Review* 85 (2): 483–513.

Higgins, E., and J. Mason. 2004. What is the value of recourse to asset backed securities? A study of credit card bank ABS rescues. *Journal of Banking & Finance* 28 (4): 857–874.

Hong, H., T. Lim, and J. Stein. 2000. Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies. *The Journal of Finance* 55 (1): 265–295.

Keynes, J. 1937. The general theory of employment. *The Quarterly Journal of Economics* 51 (2): 209–223.

Keys, B., T. Mukherjee, A. Seru, and V. Vig. 2009. Financial regulation and securitization: Evidence from subprime loans. *Journal of Monetary Economics* 56 (5): 700–720.

Klein, R., and V. Bawa. 1976. The effect of estimation risk on optimal portfolio choice. *Journal of Financial Economics* 3 (3): 215–231.

Knight, F. 1921. *Risk, Uncertainty and Profit*. Boston, MA: Houghton Mifflin Company.

Krishnaswami, S., and V. Subramaniam. 1999. Information asymmetry, valuation, and the corporate spin-off decision. *Journal of Financial Economics* 53 (1): 73–112.

Landsman, W., K. Peasnell, and C. Shakespeare. 2008. Are asset securitizations sales or loans? *The Accounting Review* 83 (5): 1251–1272.

Lang, M., and R. Lundholm. 1996. Corporate disclosure policy and analyst behavior. *The Accounting Review* 71 (4): 467–492.

Leuz, C., and R. Verrecchia. 2000. The economic consequences of increased disclosure. *Journal of Accounting Research* 38 (Supplement): 91–125.

——— 2003. IAS versus U.S. GAAP: Information asymmetry-based evidence from Germany's new market. *Journal of Accounting Research* 41 (3): 445–472.

Mason, J. 2008. Cliff risk and the credit crisis. Working paper, Louisiana State University.

—— 2009. Subprime servicer reporting can do more for modification than government subsidies. Working paper, Louisiana State University.

Minton, B., R. Stulz, and R. Williamson. 2009. How much do banks use credit derivatives to hedge loans? *Journal of Financial Services Research* 35 (1): 1–31.

Niu, F., and G. Richardson. 2006. Are securitizations in-substance sales or secured borrowings: Capital market evidence. *Contemporary Accounting Research* 23 (4): 1105–1133.

Petersen, M. 2009. Estimating standard errors in finance panel data sets: comparing approaches. *Review of Financial Studies* 22 (1): 435–480.

Ryan, S. 2007. *Financial Instruments and Institutions: Accounting and Disclosure Rules*. 2nd edition. Hoboken, NJ: John Wiley & Sons, Inc.

Schipper, K., and T. Yohn. 2007. Standard-setting issues and academic research related to the accounting for financial asset transfers. *Accounting Horizons* 21 (1): 59–80.

Schwarcz, S. 2004. Rethinking the disclosure paradigm in a world of complexity. *University of Illinois Law Review* 2004 (1): 1–38.

Van Ness, B., R. Van Ness, and R. Warr. 2001. How well do adverse-selection components measure adverse selection? *Financial Management* 30 (3): 77–98.

Vermilyea, T., E. Webb, and A. Kish. 2008. Implicit recourse and credit card securitizations: what do fraud losses reveal? *Journal of Banking & Finance* 32 (7): 1198–1208.

THE ACCOUNTING REVIEW
Vol. 86, No. 2
2011
pp. 569–604
American Accounting Association
DOI: 10.2308/accr.00000021

Warranty Reserve: Contingent Liability, Information Signal, or Earnings Management Tool?

Daniel Cohen
The University of Texas at Dallas

Masako N. Darrough
Rong Huang
Baruch College–CUNY

Tzachi Zach
The Ohio State University

ABSTRACT: We examine the information role of accounting disclosures on warranties, utilizing a database that became available due to the requirements of FIN 45. First, because firms use warranty policies as a business strategy to promote their products, a warranty reserve can serve two roles: an information signal regarding product quality, as well as a contingent liability. Consistent with this view, we find that the stock market recognizes that: (1) the warranty reserve contains information about firms' future performance, and (2) the reserve is a liability. Second, because warranty accruals require estimation of future claims, they can be used as a tool of earnings management. Our evidence indicates that managers use warranty accruals to manage earnings opportunistically to meet earnings targets. Finally, we find that the stock market recognizes the understatement of warranty liabilities of firms that managed earnings.

Keywords: *warranties; contingent liability; product quality; signaling; earnings management.*

We thank Steve Kachelmeier, Wayne Thomas, and two anonymous referees for helpful comments and suggestions. We also thank Patricia Dechow (AAA discussant), Nicole Heron, participants at the 2009 AAA Annual Meeting, and seminar participants at Fordham University, The George Washington University, INSEAD, Massachusetts Institute of Technology, Tel Aviv University, Temple University, University of Rochester, Washington University in Saint Louis, the 2008 Four-School Conference at Baruch College–CUNY, and the 2008 Burton Workshop at Columbia University. Finally, we thank Eric Arnum for his help with warranty data.

Editor's note: Accepted by Wayne Thomas.

Submitted: January 2009
Accepted: August 2010
Published Online: March 2011

I. INTRODUCTION

Most durable products are sold with warranties. A product warranty is "an obligation incurred in connection with the sale of goods or services that may require further performance by the seller after the sale has taken place" (SFAS No. 5, *Accounting for Contingencies*).[1] A warranty guarantees customers that a product will provide expected service. In the event of failure, the warranty provider would rectify the product according to the terms of the warranty policy, which can vary in duration and scope (full or limited, labor and/or parts, repair versus refund, etc.). Thus, a warranty can reduce uncertainty about the product's future performance, providing insurance against failure (Heal 1977). More importantly, warranties can be a means of signaling product quality by producers when information asymmetry exists between producers and consumers (e.g., Spence 1977; Grossman 1981; Lutz 1989).

The accounting aspects of product warranties have yet to be studied. Our study helps to fill this void by investigating the role of accounting information on warranties. We use a comprehensive database of warranty disclosures that has not been available to researchers until recently. Although firms were at liberty to disclose warranty information voluntarily, Financial Interpretation No. (FIN) 45, which took effect in 2003, mandated the disclosure of such information.

We address two research questions. First, how does the capital market interpret accounting information on warranties? Specifically, we ask whether the market interprets warranty reserves as an information signal or a contingent liability.[2] Second, we ask whether firms use warranty accounting as an earnings management tool. To address these questions, we employ a sample of 806 firms that disclosed quarterly warranty information from 2003 to 2006. We supplement the data with hand-collected information on warranty durations from firms' annual reports for a subsample of 145 firms.

Our first research question examines the market valuation of warranty reserves. If firms provide warranties as insurance, then warranty liabilities are simply contingent liabilities: future obligations to perform services if a product fails. Accordingly, one dollar of warranty liabilities would be expected to reduce firm value by one dollar. However, if firms offer warranties to signal product quality, then warranty liabilities can have an additional role in providing information on firm value and future firm performance. Due to this dual nature of warranty liabilities, we expect them to differ from other liabilities such as bank loans.

Our results demonstrate that the stock market values warranty liabilities differently from other liabilities by placing a smaller negative valuation coefficient on them. However, after controlling for analysts' earnings growth expectations and the duration of warranties, the valuation coefficients on both warranty liabilities and other liabilities approach -1. We also show that the valuation coefficient increases as firms provide better warranty coverage. However, within a group of firms with the same coverage, the valuation coefficient on warranty liabilities is negative. In addition, we document a significant positive relation between abnormal warranty expense and future firm performance across firms with different warranty terms, and a negative relation within groups of firms that provide the same warranty terms. Collectively, these findings suggest that warranty liabilities are not merely a tool of insurance; rather, they reflect a signal about firm types and future performance. Moreover, the market appears to understand this information role of warranty liabilities.

Our second research question examines whether managers use warranty accounting as a means of opportunistic earnings management to meet short-term financial reporting objectives.

---

1 Warranties can be either "express" or "implied." An express warranty typically contains a written warranty policy detailing the terms of warranty. An implied warranty is an implicit understanding that the product meets the warranty of merchantability, i.e., fit for sale and consumption as represented at the time of sale.

2 Throughout the study we use the terms "warranty reserve/s" and "warranty liability/ies" interchangeably.

These opportunistic accounting decisions can be achieved through changes in the assumptions and estimates underlying warranty accruals. Achieving earnings targets, such as avoiding losses, avoiding earnings decreases, and meeting or beating analysts' forecasts, has been extensively studied in the accounting literature (e.g., Burgstahler and Dichev 1997; Degeorge et al. 1999). In general, the consensus from prior research is that managers care greatly about these benchmarks and are willing to engage in costly earnings management strategies to achieve them (e.g., Brown and Caylor 2005; Graham et al. 2005). We show that firms that achieve earnings targets report significantly lower warranty expenses than their counterparts. This evidence is consistent with managers using the flexibility in the assumptions underlying the calculation of warranty expenses to achieve specific financial reporting objectives.

Our final analysis combines the valuation and earnings management aspects. We show that, after controlling for both the information role of warranty reserves and earnings management incentives, the market views warranty liabilities similarly to other liabilities. Also, firms that use warranty accruals to achieve earnings targets have a stronger negative valuation coefficient on their warranty liabilities. This suggests that investors recognize that the warranty liabilities of these firms are understated.

We contribute to the existing accounting literature in several ways. First, we extend prior research on the role of accounting information by examining how the capital market evaluates warranty information. We take advantage of the unique economic nature of warranty information, and document that warranty liabilities play a dual role: as contingent liabilities and as signals of product quality and future earnings growth. Second, in the context of earnings management, we delve deeper into warranty accruals. Thus, we are able to: (1) increase the power of the analysis, and (2) move away from models that evaluate aggregate discretionary accruals to a more specific model that allows us to capture the richness of the warranty setting. In that, we specifically answer the calls made by accounting researchers (for example, McNichols 2003) for disaggregating empirical measures of accounting choices. In so doing, we advance the literature on earnings management and shed light on the specific methods that managers use to achieve earnings targets. Third, our results imply that the mandated disclosures of FIN 45 regarding warranties are useful and relevant to capital market participants.

In Section II we provide a brief background on the economic role and accounting treatment of warranties. We also outline the theory that underlies our hypotheses, which we develop in Section III. In Section IV we describe our sample and empirical methodologies. We report results in Section V and conclude in Section VI.

II. BACKGROUND

The Economic Role of Warranties

In the U.S., issuing warranties for consumer products has its roots in the automobile industry (see Burton 1996). As the number of automobile companies grew, they started to use warranties as a means of competition. Other manufacturers then began offering warranties as a standard practice. Ambiguities in these contracts, however, presented enforcement problems. In order to achieve a uniform standard in warranty contracts, Congress passed the Magnuson-Moss Act in 1975. The Act required that warranty plans (when offered) explicitly describe the scope and duration of coverage, the means to obtain warranty services, and how various state laws on warranties are affected. In addition, warranties are to be offered before consumers buy a product. The Act states: "The disclosure requirement's purpose was to insure that a warranty would be a useful gauge—or

signal—of a product's reliability."[3]

With the advent of disclosure requirements, warranties became an increasingly important strategic mechanism for manufacturers. The economics literature argues that warranties are a means to overcome information asymmetries regarding product quality between an informed manufacturer and uninformed customers. By issuing a warranty plan that depends on an *ex post* verifiable outcome that is correlated with product quality, the manufacturer bonds herself (and the buyer protects himself) to its product quality (Grossman 1981). Spence (1977) posits that manufacturers provide warranties with better terms to signal their product quality.[4] In the signaling context, product quality refers to products' attributes that are not easily visible to buyers, rather than to observable characteristics. Sellers turn to costly warranty plans to convince buyers that their products are of better quality and their firms are of better "type."[5] Consistent with this notion, researchers in marketing found that product reliability and warranties' duration are positively related for a wide range of consumer durable products (Gerner and Bryant 1978; Wiener 1985; Kelley 1988). In an experimental setting, Boulding and Kirmani (1993) found that consumers perceive that a better warranty implies a higher quality product.[6]

Accounting for Warranties

Manufacturers who provide product warranties are required to accrue a warranty expense at the time of sale.[7] Like many other accruals, warranty expenses are estimated based on company's projections of future claims. Warranty expenses are an important selling expense and can be substantial. In our sample, the average warranty expense constitutes about 1.4 percent of sales and 11 percent of operating income.

The disclosures of warranty expenses and liabilities were voluntary until the issuance of FIN 45, *Guarantor's Accounting and Disclosure Requirement for Guarantees, Including Indirect Guar-*

---

[3] U.S. House Committee on Interstate and Foreign Commerce (1970), *Hearings on Warranties and Guaranties*, 91st Cong., 2d sess.

[4] Another view on warranties in the economics literature is that warranties facilitate risk-sharing between sellers and buyers (Heal 1977). Sellers and buyers might be aware of the failure rate (i.e., no information asymmetry about product quality), but it may be impossible to determine if a specific item is a lemon. If warranties are provided as insurance, then differences in warranty plans mainly reflect different consumers' attitudes toward risk. In addition, the terms of warranty plans might specify the conditions under which the plan is honored, thereby promoting proper use of the product. Consumers would value products with warranties more and would be willing to pay higher prices for them. Costs of servicing warranties are additional product costs, while warranty liabilities represent contingent liabilities.

[5] Although the positive relation between product quality and warranty quality is intuitively appealing, it is by no means the only theoretical prediction in signaling games. For example, Lutz (1989) derives a separating equilibrium in which high product quality is signaled with a low warranty plan and a low product price when consumers are subject to moral hazard. Under double moral hazard (both consumers and producers), the relation between warranty policy and firm type can be either positive or negative, depending on the parameter values (Cooper and Ross 1985). Gal-Or (1989) analyzes the role of a warranty in an oligopolistic market and shows that multiple equilibria can result; warranty quality/firm-type relation is positive in one equilibrium, but negative in another. Thus, in these equilibria the information content of a warranty plan regarding firm type is extremely limited.

[6] Warranties can also be used as a marketing tool. As such, warranties reflect a firm's strategy to establish or improve its reputation among its customers. If customers infer that a company providing products with better warranty coverage is more reliable than a company providing less warranty coverage, then offering a better warranty could be an effective strategy (Murthy and Djamaludin 2002). In addition, firms may use warranties to promote future sales and growth even though it is costly to do so (Menezes and Quelch 1990).

[7] Sellers sometimes offer *extended* warranties for additional fees. Under the current accounting regulation (Technical Bulletin 90-1), revenues from extended warranties are deferred and service costs are expensed as incurred. Thus, accounting information on warranties does not include information on extended warranties. While GAAP requires warranty expense to be matched against the revenues of the products to which the warranty applies, for tax purposes the deduction is allowed only when the costs of providing the warranty services are actually incurred. Accordingly, this treatment may give rise to a deferred tax asset.

antees of Indebtedness of Others, in 2002 (FASB 2002).[8] By mandating disclosures, FIN 45 expands the information made available to investors about firms' warranty accruals, claims, and liabilities. Beginning in 2003, firms provide: (1) the estimated amount of future payments under the warranty plan (warranty liabilities), (2) the accounting policy and methodology used in determining warranty liabilities, and (3) a tabular reconciliation of the changes in the warranty liability for the reporting period. This reconciliation presents the beginning balance of the warranty liability, the reductions in that liability for payments made under the warranty plan (i.e., claims), the changes in the liability for accruals (i.e., warranty expenses) related to product warranties issued during the reporting period, the changes in the liability for accruals related to preexisting warranties (including adjustments related to changes in estimates), and the ending balance of the warranty liability. Appendix A provides two examples of warranty disclosures from the financial statements of Dell and Western Digital.

Interpretation of Warranty Data: A Signaling Perspective

Possible Equilibria

We now discuss how one could interpret the accounting information on warranties (expenses, claims, and liabilities) from a signaling perspective in which a firm designs its warranty policy to signal its firm "type" (product quality). Although the direct signaling mechanism is the warranty policy itself, accounting information on the warranty plan could also reflect firm type and provide supplementary information.

Consider the three possible equilibria in a signaling game: pooling, fully separating, and partially pooling equilibria. Assume first that warranty policies are observable without noise. If a pooling equilibrium prevails, then one cannot differentiate firm type from observing identical warranty policies. But accounting information on warranties can reveal firm type; inferior quality will result in higher claims and higher warranty expenses. In this case, quality and warranty costs[9] are negatively related.[10]

Next, consider a fully separating equilibrium in which higher-quality types provide warranty policies with better coverage.[11] In this case, warranty policies signal product quality and reveal firm type. Although accounting information reflects the cost of providing the signal, it does not provide any incremental information about firm type. Finally, in the case of a partially pooling equilibrium, warranty cost information is informative about firms within each pool, but does not provide any incremental information across pools. Again, warranty policies themselves fully reveal firm type across policy pools.

In sum, if warranty policies are observed *perfectly*, then accounting information provides incremental information about firm type only within each pool, but not across pools. However, because warranty plans are likely to be observed with noise, even for a separating equilibrium, accounting information is likely to provide incremental information.[12] Therefore, we expect ac-

---

[8] FIN 45 applies to financial reports ending after December 15, 2002. Prior to FIN 45, the disclosure on warranty obligations were voluntary unless the warranty liabilities exceeded 5 percent of total liabilities. Gu (1998) documents that in that era, firms differed in their voluntary disclosure behavior with respect to warranty information.

[9] By warranty costs, we mean costs of providing warranty services. Warranty expenses are estimates of warranty costs accrued at the time of sale.

[10] One should note, however, that while firms could offer identical warranty plans, a pooling equilibrium is sustainable only if firms in the pool are similar. Since buyers will be able to infer the quality of the products by examining warranty expenses (i.e., the higher warranty expenses, the lower the product quality), a lower quality firm is likely to reduce the level of warranty plans. Thus, in the long run, better firms are more likely to offer better plans.

[11] For simplicity, we assume that warranty coverage can be characterized by its duration and scope, and that buyers are able to assign a strict preference ordering over various plan features.

[12] There are two sources of noise. First, firms typically provide only coarse information on warranty policies, such as the range of warranty duration for their products. Even though we develop a method to evaluate various features of warranty

counting information on warranties to be value-relevant regardless of the prevailing equilibrium.

How are warranty costs and firm type related? As discussed earlier, in a pooling equilibrium in which several types send the same signal, warranty costs and type are negatively related. The relation in a separating equilibrium, however, requires further discussion.

Relation between Warranty Costs and Firm Type in a Separating Equilibrium

We first illustrate how a separating equilibrium emerges. Consider a high-quality car manufacturer. While potential buyers can observe some characteristics of the car (e.g., design, paint, engine size, performance specifications, sunroof, leather seats), they cannot observe the car's reliability. Knowing that the car is reliable, the car manufacturer offers an extensive warranty to signal its quality type. Although the car manufacturer will have higher warranty costs, the superior warranty plan will increase the demand for the car, resulting in a higher selling price and/or sales volume, which will generate higher profits. Less reliable car makers, however, find it too costly to mimic the high-reliability car maker.

For such a scenario to be an equilibrium, there must be a cost structure in which the marginal cost of providing better coverage is lower for a higher type than for a lower type (referred to as the "single crossing property"). Thus, a lower type cannot mimic a higher type (self-selection). That firm type and warranty plan are positively related in equilibrium, however, does not imply that firm type and warranty *costs* are positively related. In Appendix B, we formally demonstrate that if the cost advantage of a higher type over a lower type (two adjacent types) in providing warranties is not too large, then warranty costs and type are positively related in a separating equilibrium.

In selecting a warranty plan, a higher type trades off a higher revenue generated from a better warranty plan against its higher cost. Even though the higher type has a cost advantage over the lower type in providing the same warranty plan, the higher type's cost will be higher for a better plan than for an inferior plan. If the cost advantage is not too large, however, then the warranty cost of a better plan by a higher type is greater than the warranty cost of an inferior plan by a lower type. Thus, unless the higher types have a significant technological advantage in providing warranty plans over lower types, it is more likely that warranty costs and type are positively related in equilibrium. In most industries with many producers, it is unlikely that any firm commands such a technological advantage in providing warranty services.

The relation between warranty liabilities and product quality is determined by claim rates, timing of claims (how soon products fail and how fast consumers submit claims), and the duration of warranty coverage as well as accrued warranty expenses. *Ceteris paribus*, warranty liabilities will expire more quickly as the duration decreases and more slowly as the duration increases. Thus, for firms with similar claim patterns, we expect warranty expenses and warranty liabilities to be positively correlated.

To summarize, in a separating equilibrium, we expect a positive relation between warranty expenses and firm type as well as between warranty liabilities and firm type, if the differences in the costs of providing warranties between high-type and low-type manufacturers are not too large. Further, in a partially pooling equilibrium, we expect a positive relation between costs and firm type across different pools of firms, but a negative relation within each pool.

---

plans for a specific product and to assign a score for the warranty plans, most of these features may not be easily observable. Second, since most firms sell many products, one needs information on the warranty policies and the sales levels of all products sold by the firm in order to compute an aggregate measure for each firm. However, such disaggregated data are not available.

III. HYPOTHESES DEVELOPMENT

We now develop specific hypotheses for our empirical analyses. The first two hypotheses focus on how the capital market evaluates warranty liabilities (H1), and on the relation between warranty expenses and future firm performance (H2). In these hypotheses, we view warranty policies as part of an overall business strategy and assume that firm type (product quality) is positively related to both firm value and future performance.[13] The third hypothesis (H3) relates to the accounting choices regarding warranties and whether managers use them to manage earnings to achieve targets.

Signaling Role of Warranty Information

Because of the uncertainty associated with future claims, a product warranty falls under the definition of a contingent liability. Valuation of contingent liabilities involves assumptions and estimates that are unobservable by outsiders. Studies that investigate contingent liabilities' valuations generally find that these liabilities are negatively associated with firm value.[14]

With respect to warranty liabilities, we expect that if they do not contain any information regarding product quality and firm performance, then they will be negatively related to firm value, like other liabilities. In this case, the valuation coefficient placed on warranty liabilities is expected to equal the coefficient on other liabilities. Therefore, our first hypothesis is as follows:

H1a: The valuation coefficient placed on warranty liabilities is equal to the valuation coefficient placed on other recognized liabilities.

If, however, warranty liabilities capture the warranty policies' information on product quality, and investors recognize the dual nature of warranty liabilities, then the coefficient on warranty liabilities is likely to be less negative than the coefficient on other liabilities. We predict that the coefficient on warranty liabilities will approach that of other liabilities only when we control for the signaling role of warranty liabilities.

Perusal of warranty policies of our sample firms shows that warranty coverage varies within an industry, especially in terms of duration. However, there are clusters of firms with the same warranty duration, defined as warranty pools.[15] That is, many industries in our sample exhibit partially pooling equilibria. As discussed earlier, the relation between warranty liabilities and product quality in a partially pooling equilibrium depends on whether we are concerned with firms within a pool or across pools. Formally:

H1b: Warranty liabilities increase with firm type across pools, while they decrease with firm type within a pool.

Because firms in better pools are likely to perform better in the future, we expect the market to value the signaling component of their liabilities positively. More specifically, we expect the

---

[13] Our assumption finds support in the literature. Ittner and Larcker (1998) show that product quality increases customer satisfaction, which in turn increases firm's profitability. Nagar and Rajan (2001) document a negative relation between product defects and future sales. The literature on the Balanced Scorecard also discusses the relation between future performance and product quality as one form of nonfinancial performance measures (Kaplan and Norton 1992, 1996).

[14] These include the studies of pensions (e.g., Barth 1991; Espahbodi et al. 1991; Landsman 1986), retirees' health benefits (Mittelstaedt and Warshawsky 1993), bank loan loss provisions (Petroni 1992; Wahlen 1994; Liu et al. 1997), and environmental liabilities (Barth and McNichols 1994).

[15] In this study, we use the information on duration of warranty coverage as the proxy for the quality of warranty plans. While there are other attributes to characterize warranty coverage, duration is the most common way to summarize warranty policies. For example, the Magnuson-Moss Act states: "Only when the rules of the warranty game are clarified... so that the duration of the warranty is a useful comparative gauge of product reliability … [will it be] economically rewarding for producers of consumer products to build reliability into their products." In addition, all warranty policies must specify warranty duration. Thus, most studies on warranties use duration as a summary measure for the quality of policies.

coefficients in our valuation models to increase across pools. Within each pool, however, we expect higher warranty liabilities to reduce valuation.

Similar to the valuation context, we expect the relation between warranty expenses and future performance to vary according to the set of firms we examine. Thus:

H2: Warranty expenses increase with firm type across pools, while they decrease with firm type within a pool.

Our empirical tests for examining H2 focus on the impact of unexpected increases in warranty expenses on future firm profitability.

Benchmark Beating and Warranty Accruals

Our third hypothesis concerns the relation between warranty accounting choices and managerial incentives to meet or beat earnings benchmarks. The means by which managers achieve earnings targets are numerous and broadly include accrual management and real earnings management.[16] Several studies examine specific accrual choices and find some evidence of earnings management. By limiting attention to a specific accounting choice, these studies are able to potentially increase the power of the tests.[17] McNichols (2003) emphasizes the importance of disaggregating empirical measures of accounting choices to generate a more powerful setting. The warranty context enables us to overcome some of the difficulties posed by aggregate, accrual-based measures and to directly address the call by McNichols (2003).

If firms use warranty expenses to achieve financial reporting objectives, then there should be an association between abnormal warranty expenses and variables proxying for reporting incentives. We focus on three extensively studied earnings benchmarks: (1) avoiding reporting a loss, (2) avoiding reporting an earnings decrease, and (3) meeting analysts' forecasts.

For each of the three benchmarks, we define "suspect" firms as those firms that are more likely to have used warranty expenses to meet one of the three benchmarks. Specifically, we identify firms whose pre-managed earnings numbers fall short of the target benchmark, but whose post-managed numbers exceed the targets. Abnormal warranty expenses are used to compute pre-managed earnings. Thus, we compare abnormal warranty expenses of these firms to those of a set of non-suspect firms. Our third hypothesis, stated in alternative form, is:

H3: Firms that were just able to exceed an earnings benchmark will report lower abnormal warranty expenses for that quarter compared to other firms.

IV. SAMPLE AND EMPIRICAL METHODOLOGY

Data

FIN 45 introduced new disclosures about warranty accruals, claims, and liabilities. We obtain these data for the years 2003–2006 from *Warranty Week* (http://www.warrantyweek.com). The sample firms are drawn from the set of manufacturing firms that are expected to have significant warranty expenses. We also hand-collect information about the duration of warranties from the 10-K filings of a subset of our sample firms that belong to industries with more than ten firms in our sample.

We describe our sample construction in Table 1. The original file contains 14,224 firm-quarter observations covering 889 unique firms. Of these, we eliminate 832 observations belonging to 52

[16] Another way to achieve the important benchmark of beating analysts' forecasts is by managing analysts' expectations (Matsumoto 2002).

[17] For example, Beaver et al. (2003) study the loan loss reserves in property-casualty insurance companies. Beatty et al. (2002) provide evidence that public banks reduce loan loss reserves to avoid reporting earnings declines. See also Moehrle (2002) and Dhaliwal et al. (2004).

TABLE 1

Sample Composition

	Firm-Quarters	Firms	Firm-Quarters	Firms
Full Sample				
Original file	14,224	889		
Observations without valid Compustat GVKEY information	(832)	(52)		
Observations of firms without sales information in some quarters during the sample period	(1,469)			
Observations of firms without direct information on warranty expenses and claims in some quarters during the sample period	(1,998)			
No warranty data for any quarter	(404)	(31)		
	9,521	806	9,521	806
Valuation Analysis				
Missing information on warranty liabilities, total assets, and analysts' growth forecasts	(3,653)	(125)		
	5,868	681		
Missing abnormal warranty expenses (industry model)	(903)	(50)		
	4,965	631		
Earnings Management Analysis				
Missing abnormal warranty expenses (industry model)			(3,991)	(142)
			5,530	664
Missing ΔROA_{t+1}, ΔROA_t and sales growth			(1,036)	(78)
			4,494	586
Subsample with *TERM* Information				
Full Sample			5,868	681
Observations belong to four-digit SIC code industries with less than ten firms			(2,092)	(290)
Observations without disclosure of warranty term information in 2006 10-K filings or the disclosure is ambiguous			(2,082)	(220)
Missing *BM*, *LIFE*, and *AGE*			(200)	(26)
			1,494	145

firms for which we could not obtain valid Compustat identification information. Following several additional data requirements, the number of effective firm-quarter observations with warranty data in our analyses is 9,521.[18] However, the number of observations in our tests may be smaller depending on model-specific variable requirements, some of which are listed in Table 1. We also conduct additional analyses on a subset of firms for which we obtain information about the duration of warranties (*TERM*). We require that these firms belong to industries with at least ten firms to ensure that we obtain a reliable benchmark against which to evaluate each firm's warranty term. This requirement as well as the existence of information about warranty duration, reduces the sample in these analyses to 1,494 observations spanning 145 firms.

The sample firms originate from several industries, but as manufacturing firms, they are concentrated in a number of groups. As reported in Table 2, about 70 percent of the sample firms belong to three industry groups: industrial machinery and equipment (196 firms; 24.3 percent), electronic and other electric equipment (198 firms; 24.6 percent), and instruments (165 firms; 20.5 percent). Warranty expenses in these industries range between 1.45 percent and 1.82 percent of sales. The last column in Table 2 reports the median warranty duration (in years).

Variable Definitions

In our analyses we use three proxies for quarterly abnormal warranty expenses and abnormal claims. Our first proxy is based on the seasonal change in warranty expenses or claims, adjusted for the seasonal change in sales. In calculating this proxy, we assume that the level of warranty expenses (or claims) for firm i is proportional to its sales in quarter t, i.e., $WEXP_{i,t} = \alpha_{i,t} SALES_{i,t}$, where $\alpha_{i,t} = WEXP_{i,t-4} / SALES_{i,t-4}$. Thus, abnormal warranty expenses in our time-series seasonal model (*ABWEXP*) are:

$$ABWEXP\_{TIME}_{i,t} = \frac{WEXP_{i,t} - WEXP_{i,t-4} * \dfrac{Sales_{i,t}}{Sales_{i,t-4}}}{TA_{i,t-4}}. \qquad \text{(Time-series model)}$$

We obtain observations of each variable in quarter t and use the same variable in the same quarter in the previous year (t–4) as a benchmark to account for the seasonal change in product mix. We control for growth in a firm's operations, which is one of the important determinants of warranty accruals. Marquardt and Weidman (2004) utilize a similar model in a different context. We compute the abnormal claims (*ABCLAIM*) made during a particular quarter in the same way by replacing *WEXP* with *CLAIMS*. This is a more direct measure of changes in product quality.

Our second proxy is an industry-adjusted measure based on the membership in a common two-digit SIC code group. For each quarter, we compute the mean level of the ratio of expenses (or claims) to sales, excluding the firm for which we calculate the measure. We require at least ten firms in the industry group. We consider the deviation from the industry mean as our proxy for the industry-adjusted abnormal warranty expenses. Thus, abnormal warranty expense in our industry model is:

---

[18] We delete 1,469 observations with missing sales information in some quarters and an additional 2,402 observations that do not have direct information on warranty expenses. Warranty expenses are missing due to one of several reasons, including: firms that adopted FIN 45 after 2003, firms that only provide annual disclosures, and firms that have warranty policies but do not comply with the FIN 45 requirement. Based on these categories, we conclude that most, if not all, of our sample firms do provide warranties to their products. Therefore, filling out the missing quarters with zeros would be inappropriate.

TABLE 2

Sample Composition by Industry

SIC Code (two digits)	Industry	n	n (%)	*WEXP/SALES* (%)	*CLAIM/SALES* (%)	Duration in Years (Median)
14	Mining and Quarrying Non-Metallic Minerals	1	0.12	0.020	0.133	—
15	General Building Contractors	24	2.98	0.750	0.617	—
16	Heavy Construction, Except Building	1	0.12	1.205	0.714	—
17	Construction—Special Contractors	3	0.37	0.968	0.900	—
22	Textile Mill Products	3	0.37	1.090	1.149	—
24	Lumber and Wood Products	13	1.61	3.468	3.625	—
25	Furniture and Fixtures	20	2.48	0.612	0.597	—
26	Paper and Allied Products	1	0.12	0.065	0.053	—
28	Chemical and Allied Products	20	2.49	2.593	2.154	—
29	Petroleum and Coal Products	3	0.37	0.838	0.854	—
30	Rubber and Miscellaneous Plastics Products	12	1.49	1.079	1.109	—
32	Glass, Pottery, and Related Products	2	0.25	0.220	0.376	—
33	Primary Metal Industries	5	0.62	0.492	0.498	—
34	Fabricated Metal Products	18	2.23	0.754	0.759	—
35	Industrial Machinery and Equipment	196	24.33	1.815	2.223	1.5
36	Electronic and Other Electric Equipment	198	24.58	1.449	1.397	1.5
37	Transportation Equipment	64	7.94	1.172	1.142	2.0
38	Instruments and Related Products	165	20.48	1.550	1.426	1.0
39	Miscellaneous Manufacturing Industries	11	1.36	1.177	1.012	—
48	Communications	1	0.12	0.000	4.227	—
50	Wholesale Trade—Durable Goods	8	1.00	0.389	0.459	—
51	Wholesale Trade—Nondurable Goods	1	0.12	0.648	0.648	—
55	Automotive Dealers and Service Stations	4	0.50	0.722	0.703	—
57	Retail	1	0.12	0.000	0.057	—
63	Insurance	1	0.12	0.153	0.093	—
67	Investment Offices, Holding Offices	1	0.12	0.120	0.249	—
73	Business Services	18	2.24	0.850	0.863	—

(continued on next page)

TABLE 2 (continued)

SIC Code (two digits)	Industry	n	n (%)	*WEXP/SALES* (%)	*CLAIM/SALES* (%)	Duration in Years (Median)
75	Auto Repair, Services, and Parking	1	0.12	3.394	4.009	—
80	Services—Health	1	0.12	1.219	1.203	—
87	Engineering and Management Services	3	0.37	1.461	1.706	—
99	Non-Classifiable Establishments	6	0.74	0.705	1.714	—
		806	100.00			

$$ABWEXP\_INDUSTRY_{i,t} = \frac{WEXP_{i,t}}{SALES_{i,t}} - AVERAGE\left(\frac{WEXP_{i,t}}{SALES_{i,t}}\right)_{OTHER\_FIRMS}. \qquad \text{(Industry model)}$$

Abnormal claims are defined in the same manner.

Our third proxy considers the duration of warranties in calculating industry-adjusted abnormal warranty expenses (or claims). For each industry-quarter, we classify observations into a low-, medium-, or high-term group if the warranty duration falls below, is equal to, or exceeds the industry median, respectively. We then compute the mean level of the ratio of warranty expenses (or claims) to sales for each industry-quarter-term group, excluding the firm for which we calculate this measure. Finally, we take the deviation from the industry-quarter-term mean as our proxy for abnormal warranty expenses or claims. In addition, we use abnormal gross margin (*ABGM*) as a control variable in some of our analyses defined in an identical manner as *ABWEXP* and *ABCLAIM*.

Descriptive Statistics

Table 3, Panel A provides summary statistics that describe our sample firms. We measure all variables on a quarterly basis and report statistics over the period from the first quarter of 2003 to the fourth quarter of 2006. For some of the variables, we also provide the values for S&P 500 nonfinancial firms for comparison purposes. Our sample firms are dispersed in size, but generally smaller than the S&P 500 nonfinancial firms. The average (median) market capitalization of our sample firms is \$3.2 billion (\$678 million). The interquartile range is between \$208 million in Q1 and \$2.2 billion in Q3. The average (median) quarterly sales of our sample firms is \$639 (\$112) million. The average (median) book-to-market ratio is 0.47 (0.42), compared to 0.40 (0.35) of the S&P 500 nonfinancial firms. Our sample firms' quarterly ROA is, on average, 0.8 percent. ROA before warranty expense is on average 1.2 percent.

The average (median) warranty expense is \$8.54 (\$1.16) million, comprising about 1.4 percent of sales and 1.5 percent of total expenses. However, the average (median) ratio of warranty expenses to the absolute value of net income is 54.8 percent (13.1 percent), indicating that for many of our sample firms, the effect of managing warranty expenses could be economically significant. Finally, we find that the liability for future warranty services comprises, on average, about 4.1 percent of sample firms' total liabilities.

Table 3, Panel A further shows that abnormal warranty expenses comprise about 0.016 percent of total assets (median is 0.005 percent). The industry-adjusted warranty expense is 0.088 percent of total assets (median is 0.394 percent of total assets). The average deviation of warranty expenses from its benchmarks is small, which is not surprising because, absent product quality changes or additional factors, warranty expenses are expected to stay around the benchmark level. The average (median) quarterly warranty claims is \$7.35 million (\$1.15 million). These claims constitute about 1.3 percent of current sales. Similarly, the abnormal claims center around zero, indicating that our benchmarks are reasonable proxies of expected expenses.

Table 3, Panel B reports correlations among key variables. We focus on the warranty variables. There is a negative correlation between the fraction of warranty liabilities on firms' balance sheet and firm size, measured as either market capitalization, sales, or total assets. Further, warranty liabilities and warranty expenses are positively correlated with analysts' forecasted growth, suggesting that warranty liabilities may not only proxy for the out-of-pocket future obligations, but also contain information about future performance. We investigate this in a multivariate setting in the next section.

TABLE 3

Descriptive Statistics and Correlations

Panel A: Summary Statistics

	n	Mean	STD	Q1	Median	Q3
General Variables: S&P 500 Nonfinancial Firms (from 2003 to 2006)						
MARKET CAPITALIZATION ($ million)	6,538	20,571	38,141	4,674	9,387	18,185
QUARTERLY SALES ($ million)	6,542	3,798	7,470	744	1,729	3,623
TOTAL ASSETS ($ million)	6,534	19,469	47,272	3,550	8,582	20,181
BM	6,407	0.402	0.271	0.227	0.345	0.521
ROA	6,479	0.017	0.025	0.007	0.015	0.027
General Variables: Warranty Sample Firms (from 2003 to 2006)						
MARKET CAPITALIZATION ($ million)	4,521	3,227	9,790	208	678	2,151
QUARTERLY SALES ($ million)	4,521	639	1,807	34	112	464
TOTAL ASSETS ($ million)	4,521	2,620	8,091	137	488	1,844
BM	4,521	0.466	0.268	0.274	0.417	0.603
ROA	4,517	0.008	0.053	0.001	0.013	0.025
ROA BEFORE WEXP	4,517	0.012	0.053	0.004	0.017	0.029
ANALYST_GR (%)	4,512	16.6	8.3	12.0	15.0	19.3
Warranty-Related Variables						
WEXP ($ million)	4,521	8.541	37.927	0.252	1.155	4.770
WEXP/SALES (%)	4,521	1.377	1.336	0.479	0.962	1.863
WEXP/TOTAL ASSETS (%)	4,521	0.376	0.443	0.107	0.236	0.476
WEXP/OPINCOME (%)	4,288	10.973	152.679	1.648	5.903	14.329
WEXP/ABS(NI) (%)	4,519	54.836	306.856	5.224	13.142	32.545
WEXP/TOTAL_EXP (%)	4,247	1.478	1.438	0.494	1.024	2.048
$ABWEXP\_{time}$ (%)	4,006	−0.016	0.305	−0.092	−0.005	0.066
$ABWEXP\_{industry}$ (%)	4,521	−0.088	1.320	−0.968	−0.394	0.411
WLIAB/LIAB (%)	4,512	4.144	4.267	1.429	2.824	5.447
Claims-Related Variables						
CLAIM ($ million)	4,521	7.349	32.984	0.249	1.145	4.233

(continued on next page)

Panel A: Summary Statistics

	n	Mean	STD	Q1	Median	Q3
CLAIM/SALES (%)	4,521	1.274	1.296	0.415	0.868	1.675
CLAIM/TOTAL ASSETS (%)	4,521	0.358	0.440	0.098	0.219	0.441
CLAIM/OPINCOME (%)	4,288	9.034	169.092	1.685	5.217	13.458
$ABCLAIM_{-time}$ (%)	4,031	−0.031	0.270	−0.094	−0.009	0.056
$ABCLAIM_{\_industry}$ (%)	4,521	−0.108	1.353	−0.975	−0.414	0.321

Panel B: Correlations

	MARKETCAP	*SALES*	*TOTAL ASSETS*	*BM*	*ROA*	*WEXP/SALES*	$ABWEXP_{\_time}$
MARKET CAP		**0.737**	**0.797**	**−0.160**	**0.096**	**0.002**	−0.002
SALES	**0.864**		**0.970**	**−0.049**	**0.040**	**−0.003**	0.000
TOTAL ASSETS	**0.896**	**0.958**		**−0.054**	0.021	**−0.007**	0.000
BM	**−0.311**	**−0.059**	0.015		**−0.280**	**0.001**	**0.057**
ROA	**0.297**	**0.184**	**0.077**	**−0.440**		**0.026**	0.002
WEXP/SALES	**0.701**	**0.824**	**0.784**	**−0.036**	**0.201**		0.244
$ABWEXP_{\_time}$	−0.025	−0.001	0.010	**0.082**	−0.022	**0.214**	
$ABWEXP_{\_industry}$	−0.026	0.016	−0.001	**0.064**	**0.108**	**0.901**	**0.195**
WLIAB/LIAB	**−0.225**	**−0.240**	**−0.307**	−0.012	**0.205**	**0.632**	−0.013
ANALYST_GR	**−0.275**	**−0.472**	**−0.446**	**−0.186**	0.037	**0.017**	−0.022
CLAIM/SALES	**−0.098**	**−0.089**	**−0.086**	**0.052**	**0.020**	**0.880**	0.083
$ABCLAIM_{\_time}$	−0.037	−0.022	−0.003	**0.096**	**−0.072**	**0.072**	**0.481**
$ABCLAIM_{\_industry}$	**−0.055**	−0.007	−0.012	**0.119**	0.024	**0.754**	**0.071**

	$ABWEXP_{\_industry}$	*WLIAB/LIAB*	*ANALYST_GR*	*CLAIM/SALES*	$ABCLAIM_{\_time}$	$ABCLAIM_{\_industry}$
MARKET CAP	**−0.017**	**−0.086**	**−0.112**	**−0.008**	0.006	**−0.033**
SALES	**0.008**	**−0.101**	**−0.197**	**−0.017**	0.003	**−0.009**
TOTAL ASSETS	**−0.007**	**−0.124**	**−0.188**	**−0.016**	0.009	**−0.019**
BM	**0.048**	0.008	**−0.189**	**0.046**	**0.072**	**0.090**
ROA	0.033	**0.142**	**−0.097**	**−0.064**	**−0.052**	**−0.053**

(continued on next page)

	$ABWEXP_{\_industry}$	$WLIAB/LIAB$	$ANALYST\_GR$	$CLAIM/SALES$	$ABCLAIM_{\_time}$	$ABCLAIM_{\_industry}$
$WEXP/SALES$	**0.940**	**0.539**	0.033	**0.861**	**0.073**	**0.793**
$ABWEXP_{\_time}$	**0.234**	0.002	−0.034	**0.083**	**0.450**	**0.074**
$ABWEXP_{\_industry}$		**0.507**	**0.028**	**0.805**	**0.076**	**0.850**
$WLIAB/LIAB$	**0.570**		**0.118**	**0.518**	−0.006	**0.480**
$ANALYST\_GR$	−0.005	**0.148**		**0.012**	−0.014	0.010
$CLAIM/SALES$	**0.787**	**0.612**	**−0.022**		**0.206**	**0.933**
$ABCLAIM_{\_time}$	**0.074**	−0.018	−0.006	**0.173**		**0.205**
$ABCLAIM_{\_industry}$	**0.864**	**0.520**	**−0.047**	**0.859**	**0.166**	

Spearman correlations are reported on the lower left and Pearson correlations are reported on the upper right. Significance level at the 5 percent level is depicted with bold font.
All variables are calculated at the end of each fiscal quarter.

Variable Definitions:
MARKET CAP = quarterly closing price multiplied by number of common shares outstanding;
SALES = quarterly sales revenue;
TOTAL ASSETS = total assets measured at the end of fiscal quarter;
BM = book value of equity divided by market value of equity;
ROA = (income before extraordinary items$_t$ + warranty expense$_t$) / $TOTAL\ ASSETS_{t-1}$;
WEXP = warranty expense;
ABWEXP = abnormal warranty expense based on either the time-series model or the industry model;
WLIAB = warranty liability;
ANALYST_GR = analyst long-term earnings growth forecasts as reported in I/B/E/S;
CLAIM = claim costs; and
ABCLAIM = abnormal claims based on either the time-series model or the industry model.

V. RESULTS

Valuation of Warranty Liabilities

We now investigate our first hypothesis regarding how warranty liabilities are related to firm's equity market prices. Our empirical specifications are derived from the Ohlson (1995) model,[19] using shares outstanding as the deflator, consistent with Barth and Clinch (2009).[20] Specifically, we estimate several variations of the following model for firm i and time t:

$$P_{i,t} = \beta_0 + \beta_1 ASSET_{i,t} + \beta_2 WLIAB_{i,t} + \beta_3 OTHER\_LIAB_{i,t} + \beta_4 ANALYST\_GR_{i,t} + \beta_5 NI_{i,t} + \beta_6 NI_{i,t}{*}Q_1 + \beta_7 NI_{i,t}{*}Q_2 + \beta_8 NI_{i,t}{*}Q_3 + Controls + \varepsilon_{i,t}, \quad (1)$$

where $P_{i,t}$ is stock price, $ASSET_{i,t}$ is total assets per share, $WLIAB_{i,t}$ is the warranty liability per share, $OTHER\_LIAB_{i,t}$ is total liabilities excluding the warranty liability per share, $ANALYST\_GR_{i,t}$ is analyst long-term earnings growth forecasts as reported in I/B/E/S, and $NI_{i,t}$ is earnings before extraordinary items per share of firm i in quarter t. To control for earnings seasonality, we include Q_1, Q_2, and Q_3 as indicators for the first three fiscal quarters. In addition, we augment the base model (1) by isolating the flow component of warranty expense (*WEXP*) and including additional control variables: *BM* is the book-to-market ratio, *AGE* is the age of the firm, computed as the difference between the current year and the year in which the business was originally formed. *LIFE* is the stage in the firm life-cycle constructed by Anthony and Ramesh (1992; see Table 4 for details). Excluding these control variables does not alter our main conclusions.

Table 4, Panel A reports the results of the market valuation of warranty liabilities.[21] The first two models serve as benchmarks to compare with subsequent regressions that incorporate warranty liabilities and growth expectations. Consistent with prior studies, the coefficient on book value per share (*BV*) in the first model is slightly greater than 1 (1.173) and the coefficient on earnings per share is positive and significant (15.228 in Q1, 13.972 in Q2, 14.016 in Q3, and 12.218 in Q4). When we decompose book value into assets and liabilities in the second model, the coefficient on assets is positive (0.913) and the coefficient on liabilities is negative (−0.915).

In the third model, we add the control variables and further decompose total liabilities into warranty liabilities and other liabilities. If the stock market recognizes the dual role of warranty liabilities—contingent liabilities and (positive) information signal—then we expect them to be valued less negatively than other liabilities. That is, we expect $\beta_3 < \beta_2 < 0$. We find the estimated coefficient of −0.605 on *WLIAB* is negative but insignificant at conventional levels ($t = -1.59$; $p = 0.115$). However, this coefficient is less negative than the coefficient on other liabilities, which is negative and statistically significant, rejecting H1a. Thus, the results are consistent with warranty liabilities containing an additional component beyond that of a simple liability. Furthermore, the coefficient on *WLIAB* is higher than that on *OTHER_LIAB*. We perform several additional analyses to shed light on this finding.

First, we include analysts' growth forecasts (*ANALYST_GR*) as an additional explanatory variable under the assumption that future earnings growth is positively correlated with product quality, implying that warranty liability might be priced less negatively than other liabilities. If the stock market correctly values the underlying liability part, then we expect warranty liabilities and

---

[19] This is consistent with prior research on valuation of pension liabilities (Landsman 1986; Barth 1991), liabilities of retirees' health benefits (Mittelstaedt and Warshawsky 1993), and environmental liabilities (Barth and McNichols 1994).

[20] We also deflated our variables by total assets. The results are similar to those reported in the study.

[21] In all of our regressions we base our inferences on two-tailed tests and standard errors that are clustered on both firm and fiscal period (Petersen 2009; Gow et al. 2010) to account for potential dependence across multiple observations in the panel. In addition, we winsorize the main variables at the top and bottom 1 percent.

TABLE 4
Market Valuation of Warranty Liability

Panel A: Full Sample

		Dependent Variable = $PRICE_t$							
		Coeff.	**Robust t-stat**	**Coeff.**	**Robust t-stat**	**Coeff.**	**Robust t-stat**	**Coeff.**	**Robust t-stat**
BV_t		1.173***	9.65						
$ASSET_t$	β_1			0.913***	7.35	0.945***	6.08	0.857***	6.13
$LIAB_t$				−0.915***	−4.66				
$WLIAB_t$	β_2					−0.605	−1.59	−1.014***	−2.90
$OTHER\_LIAB_t$	β_3					−0.865***	−4.66	−0.788***	−4.66
$ANALYST\_GR_t$	β_4							0.146***	2.95
$NI\ BEFORE\ WEXP_t$						16.504***	9.97	16.452***	9.83
$WEXP_t$						−16.407***	−9.86	−16.350***	−9.72
NI_t	β_5	12.218***	12.90	12.404***	12.08				
NI_Q1_t	β_6	3.010***	4.17	3.296***	4.96	8.360***	6.44	8.344***	6.46
NI_Q2_t	β_7	1.754**	2.58	1.835***	3.15	0.728	0.48	0.726	0.48
NI_Q3_t	β_8	1.798***	2.89	2.072***	5.07	3.013***	4.38	3.032***	4.38
BM_t						−20.055***	−7.08	−20.061***	−7.78
$LIFE_t$						2.469***	12.17	2.153***	8.79
AGE_t						1.132	1.38	0.872	1.04
Tests of									
$WLIAB_t = OTHER\_LIAB_t$						F = 5.49**	p = 0.02	F = 1.73	p = 0.19
$WLIAB_t = -1$						F = 10.03***	p < 0.01	F = 1.18	p = 0.28
Adj. R^2		85.8%		85.8%		90.1%		92.1%	
n		5,868		5,868		4,877		4,877	

(continued on next page)

Panel B: Subsample with *TERM* information

		Dependent Variable = $PRICE_t$			
		Coeff.	Robust t-stat	Coeff.	Robust t-stat
$ASSET_t$	β_1	1.150***	9.39	1.175***	9.51
$WLIAB_t$	β_2	−1.169**	−2.55	−1.200***	−3.53
$OTHER\_LIAB_t$	β_3	−1.011***	−8.77	−1.023***	−9.39
$ANALYST\_GR_t$	β_4			0.210***	8.09
$SQRT\_TERM$	β_5	2.342***	3.23	1.440**	2.59
$NI\ BEFORE\ WEXP_t$	β_6	18.667***	15.08	18.636***	15.18
$WEXPt$	β_7	−18.467**	−14.84	−18.442***	−14.94
NI_Q1_t	β_8	2.264	1.39	2.550	1.61
NI_Q2_t	β_9	1.094	0.70	1.508	0.99
NI_Q3_t	β_{10}	0.636	0.49	0.793	0.62
BM_t		−26.520***	−10.58	−25.819***	−10.50
$LIFE_t$		1.533***	8.77	1.303***	7.55
AGE_t		0.582	1.23	0.847*	1.84
Tests of					
$WLIAB_t = OTHER\_LIAB_t$		F = 0.57	p = 0.45	F = 0.49	p = 0.48
$WLIAB_t = -1$		F = 0.19	p = 0.66	F = 0.28	p = 0.60
Adj. R^2		90.4%		90.8%	
n		1,494		1,494	

Panel C: Subsample with *TERM* Information (Alternative Specification)

		Coeff.	Robust t-stat
$ASSET_t$	β_1	1.092***	10.56
$WLIAB_t$	β_2	−1.807***	−7.36
$WLIAB_t * TERM\_MEDIUM$	β_3	0.774***	4.71
$WLIAB_t * TERM\_HIGH$	β_4	1.379***	6.66
$OTHER\_LIAB_t$	β_5	−1.005***	−8.20

(continued on next page)

Panel C: Subsample with *TERM* Information (Alternative Specification)

		Coeff.	Robust t-stat
$NI\ BEFORE\ WEXP_t$	β_6	14.585***	10.12
$WEXP_t$	β_7	−13.937***	−9.01
NI_Q1_t	β_8	2.013	1.27
NI_Q2_t	β_9	−0.575	−0.38
NI_Q3_t	β_{10}	−1.114	−1.09
BM_t		−24.881***	−10.53
$LIFE_t$		1.504***	10.84
AGE_t		1.254***	2.70
$TERM\_MEDIUM$		1.198***	3.66
$TERM\_HIGH$		5.937***	6.01
Tests of			
$WLIAB_t = OTHER\_LIAB_t$		$F = 6.09$***	$p = 0.01$
$WLIAB_t * (1 + TERM\_MEDIUM) = OTHER\_LIAB_t$		$F = 0.29$	$p = 0.59$
$WLIAB_t * (1 + TERM\_HIGH) = OTHER\_LIAB_t$		$F = 5.88$**	$p = 0.02$
$WLIAB_t = -1$		$F = 5.54$**	$p = 0.02$
$WLIAB_t * (1 + TERM\_MEDIUM) = -1$		$F = 0.83$	$p = 0.36$
$WLIAB_t * (1 + TERM\_HIGH) = -1$		$F = 3.96$**	$p = 0.05$
$WLIAB_t < 0$		$F = 20.55$***	$p < 0.01$
$WLIAB_t * (1 + TERM\_MEDIUM) < 0$		$F = 12.38$***	$p < 0.01$
$WLIAB_t * (1 + TERM\_HIGH) < 0$		$F = 7.89$***	$p < 0.01$
$WLIAB_t * TERM\_MEDIUM = WLIAB_t * TERM\_HIGH$		$F = 4.08$**	$P = 0.04$
Adj. R^2		94.9%	
n		1,494	

*, **, *** Signify statistical significance of two-tailed tests at 10 percent, 5 percent, and 1 percent, respectively.
All panels of Table 4 examine the market valuation of warranty liabilities. The dependent variable is price per share. Coefficients on industry (two-digit SIC code) and quarterly indicators are not shown. Robust t-statistics are based on standard errors that are clustered by both firm and quarter.

(continued on next page)

Variable Definitions:

BV = book value per share;
ASSET = total assets per share;
LIAB = total liabilities per share;
WLIAB = warranty liabilities per share;
OTHER_LIAB = total liabilities excluding the warranty liability per share;
NI = earnings before extraordinary items per share;
SQRT_TERM = square root of *TERM* where *TERM* is defined as 1 if the warranty duration is below industry median, 2 if it equals industry median, and 3 if it is above industry median, where industry is defined at the four-digit SIC level with at least ten firms in each industry;
ANALYST_GR = analyst long-term earnings growth forecasts as reported in I/B/E/S; *Q1, Q2, Q3* are indicators for fiscal quarter 1, 2, and 3, respectively;
BM = book-to-market ratio;
AGE = age of the firm computed as the difference between the current year and the year in which the business was originally formed;
LIFE = stage in the firm life cycle defined as a composite index based on yearly rankings of the following variables: (1) median annual dividend as a percentage of income for the prior five years, (2) median annual percent sales growth for the prior five years, (3) median capital expenditure as a percentage of total market value of the firm, and (4) age of the firm (Anthony and Ramesh 1992);
TERM_MEDIUM = indicator variable with the value of 1 when the firm's warranty coverage equals to industry median, and 0 otherwise; and
TERM_HIGH = indicator variable with the value of 1 when the firm's warranty coverage is greater than industry median, and 0 otherwise.

other liabilities to be valued similarly after controlling for growth. The results in Table 4, Panel A indicate that *ANALYST_GR* is positively related to equity prices ($t = 2.95$; $p < 0.01$). Moreover, by including this variable, the coefficient of -1.014 on *WLIAB* becomes significantly negative and close to -1 ($t = -2.90$; $p < 0.01$). An F-test does not reject the hypothesis that the coefficient on *WLIAB* is equal to that on other liabilities ($p = 0.19$). A second F-test does not reject the hypothesis that the coefficient on *WLIAB* is equal to -1 ($p = 0.28$). Combined, these findings are consistent with the market recognizing the signaling component of warranty liabilities, as captured by analysts' growth forecasts.

Second, to further investigate the information role of warranty liabilities, we examine the relation between product quality and warranty liabilities, controlling for the direct signal used by firms. We add to our analysis the terms of warranties' coverages issued by the sample firms as reflected by the warranties' durations (*TERM*). We define *TERM* to equal: (i) 1, if a firm's warranty duration is lower than the industry's median, (ii) 2, if a firm's warranty duration is equal to the industry's median, and (iii) 3, if a firm's warranty duration is higher than the industry's median. We define *TERM* as a relative variable because durations of warranty policies are related to the nature of the products and the industry. Therefore, in a cross-sectional test the relative duration of warranty policies is more informative than their absolute number. We estimate the following model on a subset of firms and report results in Table 4, Panel B:[22]

$$\begin{aligned} P_{i,t} = {} & \beta_0 + \beta_1 ASSET_{i,t} + \beta_2 WLIAB_{i,t} + \beta_3 OTHER\_LIAB_{i,t} + \beta_4 ANALYST\_GR_{i,t} \\ & + \beta_5 SQRT\_TERM + \beta_6 NI\_BEFORE\_WEXP + \beta_7 WEXP + \beta_8 NI_{i,t}{*}Q_1 + \beta_9 NI_{i,t}{*}Q_2 \\ & + \beta_{10} NI_{i,t}{*}Q_3 + Controls + \varepsilon_{i,t}. \end{aligned} \tag{2}$$

In the first model, we introduce *SQRT_TERM*, the square root of *TERM*, to control for warranties' duration, allowing for nonlinearities in this relation.[23] The results show that *SQRT-_TERM* is positively related to market values ($t = 3.23$; $p < 0.01$). That is, firms that issue longer-term warranties garner a higher stock price. The square-root specification assumes that the marginal benefit of longer-term warranty diminishes. The coefficient of -1.169 on *WLIAB* ($t = -2.55$; $p = 0.01$) is not statistically different from -1 ($p = 0.66$). In the second model of Panel B, we include *SQRT_TERM,* and *ANALYST_GR*. As in the previous model, the coefficient on *WLIAB* is not statistically different from -1. *SQRT_TERM* remains significant and *ANALYST_GR* is positive and significant ($t = 8.09$; $p < 0.01$). This is consistent with *TERM* serving as a harbinger of the portions of future higher growth that are not captured by analysts' forecasts.

Third, we control for *TERM* in an alternative way to test H1b. This specification allows us to estimate the incremental valuation effect of warranty policies both across as well as within warranty pools. We define three pools and use *TERM_MEDIUM* and *TERM_HIGH* as indicator variables that equal 1 if the warranty terms are equal to or higher than the industry median. We then interact these variables with *WLIAB*. We estimate the following model:

$$\begin{aligned} P_{i,t} = {} & \beta_0 + \beta_1 ASSET_{i,t} + \beta_2 WLIAB_{i,t} + \beta_3 {*} WLIAB_{i,t}{*}TERM\_MEDIUM_{i,t} \\ & + \beta_4 {*} WLIAB_{i,t}{*}TERM\_HIGH_{i,t} + \beta_5 OTHER\_LIAB_{i,t} + \beta_6 NI\ BEFORE\ WEXP_{i,t} \\ & + \beta_7 WEXP_{i,t} + \beta_8 NI_{i,t}{*}Q_1 + \beta_9 NI_{i,t}{*}Q_2 + \beta_{10} NI_{i,t}{*}Q_3 + Controls + \varepsilon_{i,t}. \end{aligned} \tag{3}$$

---

[22] In untabulated results we estimate all models in Table 4, Panel A on the subsample of firms with *TERM* information. The inferences are very similar to those in Panel A.

[23] A square root transformation is a special case of the Box-Cox family of transformations. Because the values assigned to the variable *TERM* are essentially rankings, square roots represent an increasing and mildly concave function. We obtain similar results with several other specifications.

Based on H1b, we expect the coefficients to be negative within each pool, but to increase across pools: i.e., $\beta_2 < 0$, $\beta_2 + \beta_3 < 0$, $\beta_2 + \beta_4 < 0$ (within each pool) and $\beta_2 < \beta_2 + \beta_3 < \beta_2 + \beta_4$ (across pools). The results in Table 4, Panel C are consistent with H1b. First, they show that within each pool, *WLIAB* and firm value are negatively related. The coefficient of -1.807 on *WLIAB* of the lowest pool (β_2) is negative and significant ($t = -7.36$; $p < 0.01$). For the *MEDIUM* pool, the coefficient ($\beta_2 + \beta_3$) is -1.033 and for the *HIGH* pool ($\beta_2 + \beta_4$), it is -0.428; both are significantly different from zero ($p < 0.01$).[24] Second, across pools, we observe that the coefficients are positive and increasing. That is, β_3 is 0.774 and significant ($t = 4.71$; $p < 0.01$) indicating that $\beta_2 < \beta_2 + \beta_3$, and β_4 is 1.379 and significant ($t = 6.66$; $p < 0.01$). Third, β_4 is larger than β_3, as indicated by the last F-test reported in Panel C ($F = 4.08$; $p = 0.04$). Fourth, the coefficient on *WLIAB*, β_2, which represents the coefficient for the lowest warranty pool is significantly lower than -1, suggesting that these firms are penalized by the market. Finally, consistent with the model in Equation (2), there is some evidence that the marginal benefit of offering a higher level of warranties declines ($0.774 > 1.379 - 0.774 = 0.605$; F-stat $= 3.45$, $p = 0.06$).

Association of Warranty Information with Future Performance

Another way to shed light on the relation between warranty liabilities and equity values is to examine whether specific components of warranty expenses, in particular abnormal warranty expenses, serve as an indicator of future firm performance. An observed positive relation with future performance could explain the association between warranty liabilities and market values, which we document in Table 4. Specifically, to test H2, we estimate the following model:

$$\Delta ROA_{i,t+\tau} = \beta_0 + \beta_1 ABWEXP_{i,t} + \beta_2 ABCLAIM_{i,t} + \beta_3 ABGM_{i,t} + \beta_4 SALES\_GR_{i,t} + \beta_5 \Delta ROA_{i,t} + \beta_6 SIZE_{i,t} + \beta_7 BM_{i.t} + \varepsilon_{i,t}, \quad (4)$$

where $\tau = 1,2$.[25] We define *ROA* as earnings before warranty expenses and extraordinary items, to avoid any mechanical relation between warranty expenses and future ROA. To ensure that our estimation is robust to the model chosen, we perform the analysis using both the time-series and the two industry models. We include additional controls as follows. Abnormal warranty claims (*ABCLAIM*) controls for changes in product quality. We expect it to have a negative coefficient since higher claim costs are likely to lead to poor future firm performance (Nagar and Rajan 2001). Abnormal gross margin (*ABGM*) may also control for product quality as firms providing high-quality products are able to extract higher margins from their customers. We expect a positive coefficient on this variable because high-quality firms are generally more profitable. We expect both current sales growth (*SALES_GR*) and current change in ROA (ΔROA) to be positively related to the dependent variables because these variables persist in the short run. We expect the coefficient on *BM* to be negative because it is negatively correlated with growth opportunities. We do not make any prediction on the signs of *SIZE*.

Table 5 reports the results of estimating model (4) across pools of firms, providing a test of H2. The first and fourth columns present results using the time-series-based measures of abnormal warranty expenses ($ABWEXP\_{TIME}$) and abnormal claims ($ABCLAIM\_{TIME}$) as independent variables. We find that abnormal warranty expenses are positively associated with changes in ROA in

---

24 We also estimated Equation (3) for each industry at the two-digit SIC code level, using warranty duration to define each pool. The results are qualitatively identical to the results reported in Table 4, Panel C.

25 We also estimate the model with sales growth, *SALES_GR*, as the dependent variable. The tenor of the results remains unchanged. In addition, when we include as dependent variables the accounting-based metrics in quarter $t+3$, results are similar to those reported for quarter $t+2$.

TABLE 5

Future Performance and Abnormal Warranty Expense

	Dependent Variables					
	ΔROA_{t+1}			ΔROA_{t+2}		
	Time-Series Model	Industry Model	Industry Model Controlling for *TERM*	Time-Series Model	Industry Model	Industry Model Controlling for *TERM*
	Coefficient (Robust t-statistic)	Coefficient (Robust t-statistic)	Coefficient (Robust t-statistic)	Coefficient (Robust t-statistic)	Coefficient (Robust t-statistic)	Coefficient (Robust t-statistic)
Intercept	0.041	−0.744*	−0.446**	0.304	−0.980*	−0.338
	(0.09)	(−1.80)	(−2.40)	(1.33)	(−1.68)	(−1.49)
$ABWEXP_t$	0.734***	0.372***	0.383***	0.701*	0.189*	0.264*
	(2.77)	(3.02)	(2.80)	(1.77)	(1.96)	(1.87)
$ABCLAIM_t$	−0.938***	−0.290*	−0.168	−1.094***	−0.083	−0.083
	(−4.37)	(−1.74)	(−1.25)	(−2.88)	(−0.79)	(−0.59)
$ABGM_t$	0.327***	0.002	0.001*	0.128**	0.003	0.000
	(2.75)	(0.83)	(1.86)	(2.10)	(1.39)	(0.64)
$SALES\_GR_t$	0.017***	0.013***	0.015***	0.009**	0.009***	0.012*
	(3.32)	(4.34)	(3.60)	(2.62)	(3.35)	(1.87)
ΔROA_t	0.231***	0.628***	0.537***	0.132***	0.541***	0.595***
	(4.03)	(13.69)	(7.18)	(4.65)	(10.36)	(6.19)
$STD\ (OI/SALES)_t$	−0.862			−0.947		
	(−0.13)			(−0.17)		
$SIZE_t$	−0.002	0.239***	0.400***	−0.009	0.280***	0.533***
	(−0.05)	(4.79)	(3.02)	(−0.08)	(4.15)	(3.09)
BM_t	−1.271***	−1.685***	−1.861**	−0.702***	−1.692***	−1.132
	(−2.69)	(−4.45)	(−2.32)	(−2.77)	(−4.05)	(−1.03)
Adj. R^2	12.0%	34.7%	27.2%	5.5%	24.5%	23.8%
n	3,974	4,494	1,568	3,476	4,029	1,388

*, **, *** Signify statistical significance of two-tailed tests at 10 percent, 5 percent, and 1 percent, respectively.
ΔROA, *SALES_GR*, *ABWEXP*, *ABCLAIM*, and *ABGM* are expressed in percentages. In the industry model, all variables are measured as the deviation from the industry average of other firms where the industry is defined at the two-digit SIC level with at least ten firms in each industry. The robust t-statistics are based on standard errors that are clustered by both firm and quarter. Coefficients on industry and quarterly indicators are not shown.

Variable Definitions:

ROA = defined as earnings before extraordinary items and warranty expenses deflated by beginning-of-year total assets; and

STD (OI/SALE) = standard deviation of operating income deflated by sales for the past 8 quarters.

the next quarter (t = 2.77; $p < 0.01$) and marginally significant in quarter $t+2$ (t = 1.77; p = 0.08). This positive relation is consistent with H2 and the information role of warranties across pools of firms. The sign on *ABCLAIM*, which tracks changes in product quality, is negative and significant (coef. = -0.938, t = −4.37; $p < 0.01$). This finding is consistent with the ability of changes in product quality, as reflected in abnormal claims, to predict future firm performance

(Nagar and Rajan 2001). The results for the industry models are similar in terms of the sign, but differ in the magnitude of the coefficients because of the different deflator used for the time-series model (total assets) compared to the industry model (sales).

Overall, we conclude that abnormal warranty expenses contain information that is positively related to future firm performance. This relation, in turn, is consistent with warranty liabilities' positive association with market values as documented in the previous section. The source of information in abnormal expenses could be a result of managers applying "informative" discretion to reported earnings by, for example, altering assumptions about expected future failure rates (e.g., Watts and Zimmerman 1986; Bernard and Skinner 1996; Subramanyam 1996).

The cross-sectional analysis in Table 5 covers many firms with different warranty terms across several industries. Thus, it provides evidence on the overall association of abnormal warranty expenses and future performance (H2). However, it is also interesting to separately investigate the association between abnormal warranty expenses and future performance in groups of firms that are in the same warranty pool. In this setting, we expect warranty expenses and future performance to be negatively related, as predicted in H2. In Table 6, we estimate the same model as in Table 5 in four different pools, with each pool consisting of firms in the same industry with the same warranty term. Since this analysis requires a sufficient number of observations in the same industry and warranty pool, we estimate it in four industries. Our focus is on the coefficient on *ABWEXP*. When the dependent variable is changes in ROA in the next quarter, the coefficient is negative and significant in three of the four pools. The results are slightly weaker when we use the changes in ROA in quarter $t+2$. Overall, the analysis provides some support for H2: warranty expenses and future performance are negatively related within pools.

In sum, these analyses demonstrate that warranty liabilities serve two roles, first as a contingent liability and second as an information signal. We show that without controlling for their information role, it is not possible to assess the relation between warranty liabilities and firm value accurately. Further, we show a positive overall association between future firm performance and both warranty liabilities and abnormal warranty expense. This relation, however, is negative in groups of firms that provide identical warranty terms.

Benchmark Beating and Warranty Expenses

In this section, we test our third hypothesis regarding the relation between abnormal warranty expenses and short-term incentives to meet or beat financial reporting benchmarks. We estimate the following regression model:

$$Y_{i,t} = \beta_0 + \beta_1 SUSPECT_{i,t} + \beta_2 ABCLAIM_{i,t} + \beta_3 ABGM_{i,t} + \beta_4 BENCHMARK_{i,t} + \beta_5 SIZE_{i,t} + \beta_6 BM_{i,t} + \varepsilon_{i,t}. \tag{5}$$

The dependent variable, Y, is abnormal warranty expenses based on either the time-series or the two industry models.[26] The explanatory variable of interest is *SUSPECT*, defined in three alternative ways: (1) *SUSPECT_ΔNI* takes the value of 1 if the change in pre-managed net income is negative and the change in reported net income is positive, where pre-managed net income is defined as net income before abnormal warranty expense; (2) *SUSPECT_NI* takes the value of 1 if pre-managed net income is negative and reported net income is positive; (3) *SUSPECT_MEET* takes the value of 1 if pre-managed earnings per share misses the last outstanding analyst consensus forecast prior to the quarterly earnings announcement, while the reported earnings per share

[26] It is important to note that the dependent variable, abnormal warranty expenses (*ABWEXP*), contains some measurement error. However, because we do not believe that there is a correlation between the measurement error and our independent variables, the reported results are not biased. Instead, our model will experience a reduction in explanatory power.

TABLE 6

Future Performance and Abnormal Warranty Expense
(subsample of firms with homogenous warranty plans in different industries)

	Pool 1		Pool 2		Pool 3		Pool 4	
	ΔROA_{t+1}	ΔROA_{t+2}	ΔROA_{t+1}	ΔROA_{t+2}	ΔROA_{t+1}	ΔROA_{t+2}	ΔROA_{t+1}	ΔROA_{t+2}
	Coefficient (Robust t-statistic)	Coefficient (Robust t-statistic)	Coefficient (Robust t-statistic)	Coefficient (Robust t-statistic)	Coefficient (Robust t-statistic)	Coefficient (Robust t-statistic)	Coefficient (Robust t-statistic)	Coefficient (Robust t-statistic)
Intercept	1.099*	−1.381	1.220	1.441**	−0.420	1.093***	−0.522	0.841**
	(1.69)	(−0.60)	(0.59)	(2.25)	(−1.50)	(3.88)	(−0.59)	(2.06)
$ABWEXP_t$	−2.590***	−0.231	−1.335***	−0.742	−0.359*	−0.218	−0.892**	−0.609*
	(−3.36)	(−0.22)	(−2.90)	(−1.53)	(−1.69)	(−0.92)	(−2.24)	(−1.77)
$ABCLAIM_t$	−0.785**	−0.035	−3.226***	−0.512	−0.496*	−0.247	−0.633**	−0.383
	(−2.38)	(−0.09)	(−3.04)	(−1.49)	(−1.93)	(−0.57)	(−2.26)	(−0.71)
$ABGM_t$	0.301***	0.054	0.428*	0.059*	0.069***	0.099***	0.072	0.034
	(3.19)	(0.35)	(1.67)	(1.74)	(3.53)	(3.76)	(1.12)	(0.53)
$SALES\_GR_t$	0.026**	0.029**	0.055***	0.007	0.012	0.009	0.074**	0.004
	(2.66)	(2.17)	(3.73)	(0.39)	(1.07)	(0.81)	(2.45)	(0.37)
ΔROA_t	0.955***	1.042***	0.351	0.427**	0.674***	0.022	0.329	0.127*
	(4.06)	(3.38)	(0.72)	(2.00)	(3.78)	(0.46)	(1.23)	(1.66)
$SIZE_t$	1.152**	0.031	3.104**	0.047	0.254	0.422**	0.117	0.058
	(2.07)	(0.05)	(2.62)	(0.26)	(0.56)	(2.18)	(0.36)	(0.30)
BM_t	−3.612**	−0.977	−1.600	−1.856	−0.140	−0.081	−2.434	−3.163***
	(−2.21)	(−1.42)	(−1.28)	(−0.71)	(−0.09)	(−0.08)	(−1.14)	(−4.03)
Adj. R^2	68.7%	53.7%	59.6%	48.4%	73.0%	44.4%	70.8%	58.6%
n	100	85	93	79	71	54	61	44

*, **, *** Signify statistical significance of two-tailed tests at 10 percent, 5 percent, and 1 percent, respectively.

(continued on next page)

TABLE 6 (continued)

All variables are measured as the deviation from the average of other firms in the same pool. A pool is defined as a group of firms with the same four-digit SIC code and the same warranty duration. Pools with less than ten firms are removed from the sample. Pool 1 includes firms with four-digit SIC code of 3845 and warranty duration of 1 year. Pool 2 includes firms with four-digit SIC code of 3826 and warranty duration of 1 year. Pool 3 includes firms with four-digit SIC code of 3663 and warranty duration of 1 year. Pool 4 includes firms with four-digit SIC code of 3559 and warranty duration of 1 year.

meets or beats analyst consensus forecast, where pre-managed earnings per share is defined as earnings per share before abnormal warranty expense per share.[27] The other explanatory variables (*ABCLAIM* and *ABGM*) are adjusted based on either the time-series or industry models, corresponding to the adjustment of the dependent variable.

Table 7 reports the results. Under all specifications, we find strong evidence of unusually low abnormal warranty expenses in the three samples of firms that are suspected to have managed earnings to achieve benchmarks. All of the coefficients on *SUSPECT_ΔNI,* *SUSPECT_NI,* and *SUSPECT_MEET* are negative and statistically significant at conventional levels. Specifically, firms reporting an increase in reported net income have lower abnormal warranty expenses, as reflected in the statistically significant negative coefficient on *SUSPECT_ΔNI*, ranging from -0.173 ($t = -12.28$; $p < 0.01$) to -0.483 ($t = -6.27$; $p < 0.01$). This indicates that firms that are suspected to have engaged in opportunistic earnings management reduce warranty expenses significantly more than other firms. Also, the coefficients on *SUSPECT_NI* are negative and significant at the $p < 0.01$ level (ranging from -0.216, $t = -12.40$ to -0.918, $t = -8.02$). Finally, the coefficients on *SUSPECT_MEET* are significantly negative at the $p < 0.01$ level, ranging from -0.152 ($t = -15.88$) to -0.831 ($t = -9.58$).

The results in Table 7 also show that not all of the abnormal warranty expenses are attributable to earnings management. The consistently positive coefficient on *ABCLAIM* in all three benchmark specifications (both in the time-series and in the industry-adjusted model) suggests that as the amount of claims increases, firms allocate more warranty expenses. Overall, the results are consistent with managers using the flexibility in assumptions underlying the warranty expense calculation and exercising their discretion to achieve financial reporting benchmarks.

Valuation of Warranty Liability Combining Growth Expectation and Earnings Management Incentives

Finally, we investigate the market valuation of warranty liabilities by incorporating their contingent liability element, their information signaling role, and short-term earnings management incentives. We use an extension of model (1) as follows:

$$\begin{aligned} P_{i,t} = {} & \beta_0 + \beta_1 ASSET_{i,t} + \beta_2 WLIAB_{i,t} + \beta_3 OTHER\_LIAB_{i,t} + \beta_4 SUSPECT_{i,t} * WLIAB_{i,t} \\ & + \beta_5 SUSPECT_{i,t} + \beta_6 ANALYST\_GR_{i,t} * WLIAB_{i,t} + \beta_7 ANALYST\_GR_{i,t} + \beta_8 NI_{i,t} \\ & + \beta_9 NI_{i,t} * Q_1 + \beta_{10} NI_{i,t} * Q_2 + \beta_{11} NI_{i,t} * Q_3 + \varepsilon_{i,t}. \end{aligned} \tag{6}$$

As documented in the previous subsection, firms with strong incentives to meet or beat earnings benchmarks cut warranty expenses. We identify suspect firms as in Table 7. If investors correctly infer that these firms understate their warranty liabilities, then they would place a larger negative coefficient on warranty liabilities to correct for the underestimation.

Table 8 reports the results of model (6). In the first model, we find that the stock market places a more negative coefficient on warranty liabilities of firms that are suspected to have managed earnings to avoid reporting an earnings decline. The coefficient on the interaction term *SUSPECT * WLIAB* is -1.268 with a t-statistic of -2.67 ($p < 0.01$). In the next columns, we find similar results for suspect firms that seek to avoid a loss (coef.$= -1.832$; $t = -2.82$; $p < 0.01$), and those that seek to meet analysts' forecasts (coef.$= -0.606$; $t = -2.86$; $p < 0.01$).

Our measure of analysts' earnings growth expectations (*ANALYST_GR*) is positively associ-

[27] We also performed analysis using an alternative definition of *SUSPECT,* similar to Roychowdhury (2006). Under this definition, *SUSPECT* is defined based on the proximity of the reported accounting number to the desired benchmark. The tenor of the results is similar to that of the reported results.

TABLE 7

Incentives, Earnings Management and Warranty Expenses

	Dependent Variables = $ABWEXP_t$								
	Avoid Earnings Decline			Avoid Loss			Meet Analyst Forecast		
	Time-Series Model	Industry Model	Industry Model Controlling for *TERM*	Time-Series Model	Industry Model	Industry Model Controlling for *TERM*	Time-Series Model	Industry Model	Industry Model Controlling for *TERM*
	Coef. (Robust t-stat)	Coef. (Robust t-stat)	Coef. (Robust t-stat)	Coef. (Robust t-stat)	Coef. (Robust t-stat)	Coef. (Robust t-stat)	Coef. (Robust t-stat)	Coef. (Robust t-stat)	Coef. (Robust t-stat)
Intercept	0.028	0.011	0.003	−0.073	0.449***	0.250***	−0.021	0.363***	0.183***
	(1.35)	(0.27)	(0.05)	(−1.04)	(6.70)	(6.35)	(−1.36)	(6.67)	(2.89)
$SUSPECT\_\Delta NI_t$	−0.173***	−0.373***	−0.483***						
	(−12.28)	(−4.60)	(−6.27)						
$SUSPECT\_NI_t$				−0.216***	−0.918***	−0.679***			
				(−12.40)	(−8.02)	(−9.17)			
$SUSPECT\_MEET_t$							−0.152***	−0.831***	−0.551***
							(−15.88)	(−9.58)	(−6.86)
$ABCLAIM_t$	0.512***	0.575***	0.809***	0.446***	0.454***	0.706***	0.475***	0.489***	0.702***
	(12.79)	(6.11)	(17.66)	(9.71)	(5.57)	(15.21)	(10.37)	(7.85)	(14.87)
$ABGM_t$	0.249	0.064**	−0.072	−1.151	0.009	−0.065	−0.425	0.222	−0.423**
	(1.09)	(2.20)	(−1.42)	(−1.53)	(0.71)	(−1.50)	(−1.44)	(1.48)	(−2.23)
ΔNI_t	0.258**	5.186	3.411***						
	(2.15)	(0.93)	(4.35)						
NI_t				1.097**	2.689***	6.092***			
				(2.60)	(3.40)	(7.00)			
EPS_t							0.014**	2.163***	0.445***
							(2.50)	(2.85)	(3.90)

(continued on next page)

TABLE 7 (continued)

	Dependent Variables = $ABWEXP_t$								
	Avoid Earnings Decline			**Avoid Loss**			**Meet Analyst Forecast**		
	Time-Series Model	**Industry Model**	**Industry Model Controlling for *TERM***	**Time-Series Model**	**Industry Model**	**Industry Model Controlling for *TERM***	**Time-Series Model**	**Industry Model**	**Industry Model Controlling for *TERM***
	Coef. (Robust t-stat)	**Coef. (Robust t-stat)**	**Coef. (Robust t-stat)**	**Coef. (Robust t-stat)**	**Coef. (Robust t-stat)**	**Coef. (Robust t-stat)**	**Coef. (Robust t-stat)**	**Coef. (Robust t-stat)**	**Coef. (Robust t-stat)**
$SIZE_t$	0.008***	0.047	0.007	0.007**	0.028**	0.014	0.008***	0.002	−0.046
	(3.80)	(1.22)	(0.33)	(2.59)	(2.41)	(0.67)	(6.68)	(0.14)	(−1.60)
BM_t	0.033**	−0.106	−0.175*	0.065**	0.046	−0.139*	0.027**	0.030	−0.552***
	(2.22)	(−0.77)	(−1.83)	(2.25)	(0.78)	(−1.77)	(2.19)	(0.41)	(−3.21)
Adj. R^2	29.9%	53.0%	72.5%	31.6%	60.0%	71.1%	42.1%	66.9%	77.4%
n	4,948	5,530	1,385	5,361	6,043	1,282	3,698	4,835	1,038

*, **, *** Signify statistical significance of two-tailed tests at 10 percent, 5 percent, and 1 percent, respectively.

ABWEXP, *ABCLAIM*, and *ABGM* are expressed in percentages and are defined in Table 4. In the industry model, all variables are measured as the deviation from the industry average of other firms where the industry is defined as the two-digit SIC level with at least ten firms in each industry. The robust t-statistics are based on standard errors that are clustered by both firm and quarter. Coefficients on industry and quarterly indicators are not shown.

Variable Definitions:

SUSPECT_ΔNI = 1 if the change in pre-managed net income is negative and the change in net income is positive, where pre-managed net income is defined as net income before abnormal warranty expense;

SUSPECT_NI = 1 if pre-managed net income is negative and net income is positive;

SUSPECT_MEET = 1 if a firm's pre-managed earnings per share misses the last outstanding analyst consensus forecast prior to the quarterly earnings announcement, while the earnings per share meets or beats analyst consensus forecast, where pre-managed earnings per share is defined as earnings per share before abnormal warranty expense;

SIZE = logarithm of the market value of equity at the beginning of the quarter; and

NI = earnings before extraordinary items scaled by lagged total assets.

TABLE 8

Valuation of Warranty Liability Incorporating Growth and Earnings Management (Full Sample)

	Dependent Variable = $PRICE_t$					
	Avoid Earnings Decline		Avoid Loss		Meet Analyst Forecast	
	Coeff.	Robust t-statistic	Coeff.	Robust t-statistic	Coeff.	Robust t-statistic
$ASSET_t$	1.093***	13.22	0.944***	9.06	0.992***	9.00
$WLIAB_t$	−0.842***	−3.43	−1.082***	−4.61	−0.890***	−3.37
$OTHER\_LIAB_t$	−0.938***	−7.60	−0.769***	−5.09	−0.806***	−5.13
$SUSPECT_t * WLIAB_t$	−1.268***	−2.67	−1.832***	−2.82	−0.606***	−2.86
$SUSPECT_t$	2.841***	4.66	6.961***	8.50	4.991***	8.11
$ANALYST\_GR_t * WLIAB_t$	0.010***	3.02	0.014**	2.23	0.059***	2.76
$ANALYST\_GR_t$	0.043***	4.93	0.105***	3.93	0.118***	3.31
NI_t	13.047***	10.00	11.417***	7.23	13.450***	7.89
NI_Q1_t	3.170***	6.60	2.883***	4.19	3.119***	4.86
NI_Q2_t	1.717	1.03	1.685*	1.76	1.815*	1.68
NI_Q3_t	2.111***	4.01	1.466***	3.00	1.556***	2.90
Tests of						
$WLIAB_t * [1 + \text{Median } (SUSPECT) + \text{Median } (ANALYST\_GR_t)] = OTHER\_LIAB_t$						
	F = 0.33	p = 0.56	F = 0.46	p = 0.50	F = 1.23	p=0.27
$WLIAB_t * [1 + \text{Median } (SUSPECT) + \text{Median } (ANALYST\_GR_t)] = -1$						
	F = 0.10	p = 0.75	F = 0.59	p = 0.44	F = 2.15	p=0.14
Adj. R^2	0.873		0.880		0.894	
n	4,965		4,954		4,659	

*, **, *** Signify statistical significance of two-tailed tests at 10 percent, 5 percent, and 1 percent, respectively.

This table shows market valuation of warranty liability after incorporating earnings management incentives. The dependent variable is price per share. Coefficients on industry and quarterly indicators are not shown. All the independent variables except $SUSPECT_t$ *and* $ANALYST\_GR_t$ are deflated by common shares outstanding. The robust t-statistics are based on standard errors that are clustered by both firm and quarter.

Variable Definitions:

SUSPECT = *SUSPECT_ΔNI* in the "avoid earnings decline" regression; *SUSPECT_NI* = "avoid loss" regression; and *SUSPECT_MEET* = "meet analyst forecast" regression. *SUSPECT_ΔNI*, *SUSPECT_NI*, and *SUSPECT_MEET* are defined as in Table 7.

ated with share price across all three models. We add an interaction term between *ANALYST_GR* and *WLIAB* to examine whether the information signaling in warranty liabilities varies across firms with different growth opportunities. The coefficient on the interaction term is positive and significant for all three earnings benchmarks, indicating that warranty liabilities serve as a stronger information signal for high-growth firms than for low-growth firms.

More formally, we conduct an F-test of whether the coefficient on *OTHER_LIAB* is equal to the sum of the coefficient of *WLIAB* and its interactions with *SUSPECT* and *ANALYST_GR*, both evaluated at their median values. The results of the test indicate that there is no evidence to reject

the hypothesis that the coefficients of *WLIAB* and *OTHER_LIAB* are equal (p-values = 0.56, 0.50, and 0.27). Further, we use another F-test to examine whether the coefficient on *WLIAB* is different from -1. We cannot reject the hypothesis that $WLIAB = -1$ (p-values = 0.75, 0.44, and 0.14). In an unreported analysis of the subsample of firms for which *TERM* is available, our conclusions remain unchanged after controlling for the duration of the warranties, and incorporating duration in the industry-based measures of *ABWEXP* and *ABCLAIM*.

Overall, the results in Table 8 support the conjecture that warranty liabilities represent three aspects: a contingent liability, an information signal about growth prospects, and an earnings management tool. We find that the stock market values warranty liabilities more negatively for firms that have managed earnings and that it places a positive weight on warranty liabilities as a signal of future growth prospects. After controlling for signaling and earnings management, we find that the stock market values warranty liabilities similarly to how it values other recognized liabilities.

VI. CONCLUSION

We use a sample of over 800 firms that disclose warranty information under FIN 45 to investigate the economics and accounting aspects of product warranties. Our study provides insights into the market interpretation of warranty disclosures and managers' choices with regard to product warranty policies as well as the accounting treatment of warranties.

We first investigate the market valuation of warranty liabilities. We hypothesize that they serve as both contingent liabilities that reflect future services related to warranty obligations as well as an information signal of product quality and future growth prospects. Our findings indicate that the stock market places a smaller negative valuation coefficient on warranty liabilities compared to other reported liabilities. When we control for the signaling role of warranty liabilities (with analyst growth expectations and warranty duration), the valuation coefficients on warranty liabilities and other liabilities approach -1. We also show that the relation between market values and warranty liabilities varies across different groups of firms, depending on the level of warranty coverage they provide. The market assigns a higher valuation coefficient on warranty liabilities of firms that provide superior warranty coverage. In addition, we report a positive association between abnormal warranty expenses and future firm performance across firms that provide different warranty coverage. However, within groups of firms that provide identical coverage, this relation is negative. Together, these findings support the hypothesis that the market interprets warranty liabilities as information signals for product quality and future growth prospects.

We also find evidence that firms with incentives to manage earnings to meet earnings targets report lower abnormal warranty expenses. This evidence is consistent with managers using their discretion in the estimates of warranty accruals to achieve financial reporting targets.

Finally, we investigate the market valuation of warranty liabilities after controlling for signaling and earnings management aspects. We show that warranty liabilities reduce share prices dollar-for-dollar. We also document that investors understand that warranty liabilities of firms that engaged in earnings management are understated.

Overall, the findings in this study demonstrate a unique information aspect of warranty liabilities. Our findings also show that disclosures on warranties allow this information to be utilized by market participants.

APPENDIX A
SAMPLE WARRANTY DISCLOSURES

Dell Corp.

	Fiscal Year Ended February 1, 2008	February 2, 2007 (in millions)	February 3, 2006
Warranty liability:			
Warranty liability at beginning of year	$ 958	$ 951	$ 722
Costs *accrued* for new warranty contracts and changes in estimates for pre-existing warranties [(a)]	1,141	1,242	1,391
Service obligations honored	(1,170)	(1,235)	(1,162)
Warranty liability at end of year	$ 929	$ 958	$ 951
Current portion	$ 690	$ 768	$ 714
Non-current portion	239	190	237

(a) Changes in cost estimates related to pre-existing warranties are aggregated with accruals for new warranty contracts. Dell's warranty liability process does not differentiate between estimates made for pre-existing warranties and new warranty obligations.

Western Digital

Product Warranty Liability

Changes in the warranty accrual for 2008, 2007 and 2006 were as follows (in millions):

	2008	2007	2006
Warranty accrual, beginning of period	$ 90	$ 89	$ 92
Charges to operations	106	74	76
Utilization	(73)	(52)	(49)
Changes in estimate related to pre-existing warranties	(9)	(21)	(30)
Warranty accrual, end of period	$ 114	$ 90	$ 89

APPENDIX B

This appendix provides a brief discussion on how warranty costs and types are related in a separating equilibrium. We show that warranty costs and types are positively related if the difference in the cost of providing warranty services by type is not too large.[28]

Consider a simple model with two adjacent types: L and H. Let type H be a higher type. Both type H and L know their own type, but potential buyers do not. As the quality is higher, type H's product breaks down less frequently and is less costly to repair. Given the asymmetry of information, type H tries to signal its quality by providing a better warranty plan (w_2). Buyers believe that the provider of warranty plan 2 is the higher type. In a separating equilibrium, type H provides w_2, while type L provides w_1, which is inferior to w_2. For a product with w_2, buyers are willing to buy q_2 units at the price p_2. Type H is then expected to incur $c^H(w_2)q_2$ as the total warranty-

[28] For interested readers, a research note that provides a full discussion is available from the authors upon request.

servicing cost. Type L offers w_1, sells q_1 units at p_1, and incurs $c^L(w_1)q_1$. Thus, given the beliefs of the buyers, the warranty choices made by both types form a Perfect Bayesian equilibrium. For a separating equilibrium to exist, the higher type must have a cost advantage such that:

$$c^L(w_2) - c^L(w_1) > c^H(w_2) - c^H(w_1),\ c^L(w_1) > c^H(w_1),$$

and $c^L(w_2) > c^H(w_2)$. These inequalities are implied by the single-crossing property.

In the separating equilibrium, neither type wants to mimic the other. That is:

$$(p_1 - c^L(w_1))q_1 \geq (p_2 - c^L(w_2))q_2, \tag{B1}$$

$$(p_2 - c^H(w_2))q_2 \geq (p_1 - c^H(w_1))q_1. \tag{B2}$$

Equations (B1) and (B2) are the self-selection constraints for type L and H, respectively. Combining the two equations, we have:

$$c^L(w_2)q_2 - c^L(w_1)q_1 \geq p_2q_2 - p_1q_1 \geq c^H(w_2)q_2 - c^H(w_1)q_1. \tag{B3}$$

Equation (B3) says that the marginal benefit justifies the marginal cost of providing a better warranty plan for type H, but not for type L.

To see how total warranty costs and types are related in equilibrium, we compare $c^L(w_1)q_1$ with $c^H(w_2)q_2$. If they are to be (strictly) positively related, then we must have:

$$c^H(w_2)q_2 > c^L(w_1)q_1 \text{ or } c^H(w_2) > c^L(w_1)\frac{q_1}{q_2}. \tag{B4}$$

Condition (B4) implies that $c^H(w_2)$ cannot be too small relative to $c^L(w_1)$. While the single-crossing property requires that type H has a cost advantage, the cost advantage cannot be large.[29] For example, suppose $q_1 = q_2$ (a better warranty plan commands a higher price $[p_1 < p_2]$ but does not increase the sales volume). Then, Condition (B4) states $c^H(w_2) > c^L(w_1)$, that is, the unit cost of providing a better warranty plan (w_2) by type H is higher than that of providing the inferior plan (w_1) by type L. If $q_1 < q_2$ (a better warranty plan increases the sales volume), then it is possible to have $c^H(w_2) < c^L(w_1)$. However, $c^H(w_2)$ is bounded below.

REFERENCES

Anthony, J., and K. Ramesh. 1992. Association between accounting performance measures and stock prices: A test of the life cycle hypothesis. *Journal of Accounting and Economics* 15 (2–3): 203–227.

Barth, M. 1991. Relative measurement errors among alternative pension asset and liability measures. *The Accounting Review* 66 (3): 433–463.

———, and G. Clinch. 2009. Scale Effects in capital markets-based accounting research. *Journal of Business Finance & Accounting* 36 (3–4): 253–288.

———, and M. McNichols. 1994. Estimation and market valuation of environmental liabilities relating to superfund sites. *Journal of Accounting Research* 32 (Supplement): 177–209.

Beatty, A., B. Ke, and K. Petroni. 2002. Differential earnings management to avoid earnings declines across publicly and privately held banks. *The Accounting Review* 77 (3): 547–570.

Beaver, W., M. McNichols, and K. Nelson. 2003. Management of the loss reserve accrual and the distribution of earnings in the property-casualty insurance industry. *Journal of Accounting and Economics* 35 (3):

[29] To see that the relation can be negative, imagine a firm that produces a perfect product that never fails. The firm can signal its type by providing full warranty coverage but never incurring warranty costs. Other firms without the same technology would incur a positive cost to offer any level of warranty (or they might choose not to offer any warranty plan). The relation in this case is negative (or zero).

347–376.

Bernard, V., and D. Skinner. 1996. What motivates managers' choice of discretionary accruals? *Journal of Accounting and Economics* 22 (1-3): 313–325.

Boulding, W., and A. Kirmani. 1993. A consumer-side experimental examination of signaling theory: Do consumers perceive warranties as signals of quality. *The Journal of Consumer Research* 20 (1): 111–123.

Brown, L., and M. Caylor. 2005. A temporal analysis of quarterly earnings thresholds: Propensities and valuation consequences. *The Accounting Review* 80 (2): 423–440.

Burgstahler, D., and I. Dichev. 1997. Earnings management to avoid earnings decreases and losses. *Journal of Accounting and Economics* 24 (1): 99–126.

Burton, J. 1996. Warranty protection: A consumerist perspective. In *Product Warranty Handbook*, edited by Blischke, W. R., and D. N. P. Murthy. New York, NY: Marcel Dekker.

Cooper, R., and T. Ross. 1985. Product warranties and double moral hazard. *The Rand Journal of Economics* 16 (1): 103–113.

Degeorge, F., J. Patel, and R. Zeckhauser. 1999. Earnings management to exceed thresholds. *The Journal of Business* 72 (1): 1–33.

Dhaliwal, D., C. Gleason, and L. Mills. 2004. Last chance earnings management: Using the tax expense to achieve earnings targets. *Contemporary Accounting Research* 21 (2): 431–459.

Espahbodi, H., E. Strock, and H. Tehranian. 1991. Impact on equity prices of pronouncements related to nonpension postretirement benefits. *Journal of Accounting and Economics* 14 (4): 323–346.

Financial Accounting Standards Board (FASB). 2002. *Guarantor's Accounting and Disclosure Requirement for Guarantees, Including Indirect Guarantees of Indebtedness of Others*. FASB Interpretation No. 45. Financial Accounting Series. Norwalk, CT: FASB.

Gerner, K., and J. Bryant. 1978. The price of warranty: The case for refrigerators. *The Journal of Consumer Affairs* 12 (1): 30–47.

Graham, J., C. Harvey, and S. Rajgopal. 2005. The economic implications of corporate financial reporting. *Journal of Accounting and Economics* 40 (1–3): 3–73.

Grossman, S. 1981. The informational role of warranties and private disclosure about product quality. *The Journal of Law & Economics* 24 (3): 461–483.

Gal-Or, E. 1989. Warranties as a signal of quality. *The Canadian Journal of Economics. Revue Canadienne d'Economique* 22 (1): 50–61.

Gow, I., G. Ormazabal, and D. Taylor. 2010. Correcting for cross-sectional and time-series dependence in accounting research. *The Accounting Review* 85 (2): 483–512.

Gu, F. 1998. Voluntary disclosure of accounting information for product warranty liability. Working paper, University at Buffalo, SUNY.

Heal, G. 1977. Guarantees and risk-sharing. *The Review of Economic Studies* 44 (3): 549–560.

Ittner, C., and D. Larcker. 1998. Are nonfinancial measures leading indicators of financial performance? An analysis of customer satisfaction. *Journal of Accounting Research* 36 (Supplement): 1–35.

Kaplan, R., and D. Norton. 1992. The balanced scorecard: Measures that drive performance. *Harvard Business Review* (January-February): 71–79.

———, and ———. 1996. *The Balanced Scorecard: Translating Strategy into Action*. Boston, MA: Harvard Business School Press.

Kelley, C. 1988. An investigation of consumer product warranties as market signals of product reliability. *Journal of the Academy of Marketing Science* 16 (2): 72–78.

Landsman, W. 1986. An empirical investigation of pension fund property rights. *The Accounting Review* 61 (3): 662–691.

Liu, C. C., S. Ryan, and J. Wahlen. 1997. Differential valuation implications of loan loss provisions across banks and fiscal quarters. *The Accounting Review* 72 (1): 133–146.

Lutz, N. 1989. Warranties as signals under consumer moral hazard. *The Rand Journal of Economics* 20 (2): 239–255.

Marquardt, C., and C. Weidman. 2004. How are earnings managed? An examination of specific accruals. *Contemporary Accounting Research* 21 (2): 461–491.

Matsumoto, D. 2002. Management's incentives to avoid negative earnings surprises. *The Accounting Review*

77 (3): 483–514.

McNichols, M. 2003. Discussion of "Why are earnings kinky? An examination of the earnings management explanation." *Review of Accounting Studies* 8 (2–3): 385–391.

Menezes, M., and J. Quelch. 1990. Leverage your warranty program. *Sloan Management Review* 31 (4): 69–80.

Mittelstaedt, H. F., and M. J. Warshawsky. 1993. The impact of liabilities for retiree health benefits on share prices. *The Journal of Risk and Insurance* 60 (1): 13–35.

Moehrle, S. 2002. Do firms use restructuring charge reversals to meet earnings targets? *The Accounting Review* 77 (2): 397–414.

Murthy, D., and J. Djamaludin. 2002. New produce warranty: A literature review. *International Journal of Production Economics* 79 (3): 231–260.

Nagar, V., and M. Rajan. 2001. The revenue implications of financial and operational measures of product quality. *The Accounting Review* 76 (4): 495–513.

Ohlson, J. 1995. Earnings, book values, and dividends in security valuation. *Contemporary Accounting Research* 11 (2): 661–687.

Petersen, M. 2009. Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies* 22 (1): 435–480.

Petroni, K. 1992. Optimistic reporting in the property-casualty insurance industry. *Journal of Accounting and Economics* 15 (4): 485–508.

Roychowdhury, S. 2006. Earnings management through real activities manipulation. *Journal of Accounting and Economics* 42 (3): 335–370.

Spence, M. 1977. Consumer misperceptions, product failure and producer liability. *The Review of Economic Studies* 44 (3): 561–572.

Subramanyam, K. R. 1996. The pricing of discretionary accruals. *Journal of Accounting and Economics* 22 (1–3): 249–281.

Wahlen, J. 1994. The nature of information in commercial bank loan loss disclosures. *The Accounting Review* 69 (3): 455–478.

Watts, R., and J. Zimmerman. 1986. *Positive Accounting Theory*. Englewood Cliffs, NJ: Prentice Hall.

Wiener, J. 1985. Are warranties accurate signals of product reliability? *The Journal of Consumer Research* 12 (2): 245–250.

THE ACCOUNTING REVIEW
Vol. 86, No. 2
2011
pp. 605–635

American Accounting Association
DOI: 10.2308/accr.00000028

Investment Growth and the Relation between Equity Value, Earnings, and Equity Book Value

Shengquan Hao
Shanghai Jiaotong University

Qinglu Jin
Shanghai University of Finance and Economics

Guochang Zhang
Hong Kong University of Science and Technology

Abstract: Using Zhang (2000) as the theoretical basis, we predict and empirically test the effect of investment growth on the relation between equity value and accounting variables. We find that (1) growth increases the slope in the value-earnings relation for high-profitability firms (consistent with growth having positive NPV), but has an insignificant or negative effect on the slope for lower-profitability firms (consistent with growth having non-positive NPV); (2) given earnings, growth increases the (positive) slope of the relation between equity value and equity book value for low-profitability firms, but reduces this slope and causes equity value to *decrease* with book value for high-profitability firms; and (3) given profitability (ROE), equity value uniformly increases with book value, and growth increases the slope of this relation. We also examine the valuation impact of *past* investment activities that arises from accounting conservatism. We find that the earnings coefficient is greater in the years following faster investment increases (which cause earnings to be more conservatively stated).

Keywords: *investment growth; equity valuation; earnings; equity book value; profitability; conservatism.*

Data Availability: *The data used in this study are available from public sources.*

We thank Peter Chen for his involvement and input during early stages of this project. We also thank three anonymous referees, Gary Biddle, Bob Bowen, Kevin Chen, Zhihong Chen, Jerry Feltham, Inchi Hu, Kun Jiang, Charles Lee, Clive Lennox, Steve Penman, Terry Shevlin, Paul Zarowin (editor), Tianyu Zhang, and workshop participants at Harvard Business School, Hong Kong University of Science and Technology, Korea University, National Taiwan University, Singapore Management University, University of British Columbia, University of Toronto, and University of Washington for their comments. Professor Jin acknowledges the financial support of the MOE through the Institute of Finance and Accounting at SUFE and of the National Natural Science Foundation of China (project numbers 71032005 and 71072036). All errors are our own.

Editor's note: Accepted by Paul Zarowin.

Submitted: December 2008
Accepted: July 2010
Published Online: March 2011

I. INTRODUCTION

Growth opportunities constitute an important part of firm value (Miller and Modigliani 1961; Dixit and Pindyck 1994). When investors determine firm value based on reported accounting data, it is essential that they consider how much the firm is expected to grow. But as Holthausen and Watts (2001) observe, empirical research on accounting-based valuation has largely neglected the growth factor. In this study, we reexamine the accounting-value link by explicitly recognizing the role of growth in valuation. Using the real-options-based model of Zhang (2000) as the theoretical basis, we predict and empirically test how the relation between equity value and accounting data behaves in a distinctly nonlinear fashion and, in particular, how investment growth mediates this nonlinear relation.

In this study, growth refers to increases (or decreases) in the amount of capital invested in business operations.[1] Economic reasoning suggests that whether investment growth enhances value generation depends on profitability,[2] defined as earnings in a period divided by the equity book value at the beginning of the period (ROE). Profitability represents a firm's ability to generate value from invested capital, thus indicating the desirability of increasing or reducing the scale of operations. Mirroring its investment-guiding role, profitability thus helps investors to infer the firm's course of operations going forward. Because of its fundamental role in determining investment decisions underlying value generation, profitability (as opposed to earnings) serves as an appropriate valuation metric.[3] Obviously, investment growth undertaken by a profitable firm enhances investor value, whereas that by an unprofitable firm creates no value or even destroys value. Our analysis demonstrates how this basic notion about growth, profitability, and value creation manifests in mapping accounting data to value.

We develop four hypotheses concerning various aspects of the growth effect on the valuation function. We first test these hypotheses separately with partial model specifications (which collapses one of the dimensions of the valuation space), and then jointly with "full" model specifications. In an extended analysis, we further follow Zhang (2000), who predicts a link between the trend of past investment changes and earnings biases, to introduce the effect of conservative accounting. There, we adopt a more comprehensive empirical model within which we test the effects of both conservatism and investment growth.

Existing research shows that, given equity book value, equity value increases with earnings, and growth increases the slope of the value-earnings relation (Collins and Kothari 1989; Kothari 2001). We predict that, whereas growth strengthens the value-earnings relation in terms of a steeper slope for high-profitability firms (whose investment growth will generally have positive NPV), the effect is expected to be smaller and can even disappear for low-profitability firms (whose growth will generally have low or non-positive NPV).[4] Our empirical results are consistent. We first show that in regressions of stock price on earnings controlling for book value, the slope is positive and becomes increasingly steeper, confirming convexity as predicted by real

---

1 The growth construct in this study is input (investment) based. This is conceptually distinct from, though related to, (output-based) earnings growth. Whether investment growth translates into earnings growth depends on a firm's profitability. Ohlson and Juettner-Nauroth (2005) provide a model that relates equity value to *expected* earnings and earnings growth.

2 Fama and Miller (1972) provide an illustration of this basic notion. Bar-Yosef et al. (1987) empirically investigate the relation between investment and earnings.

3 Note that earnings cannot be compared directly across firms with different scales of operations. The importance of profitability for valuation means that income statement data should be interpreted *in conjunction with* the balance sheet, consistent with the opinion of the Financial Accounting Standards Board (see SFAC No. 5, paragraph 24a).

4 The rationale is that a firm's strength/weakness in business fundamentals should persist (to some extent) from one period to the next, so that firms with higher profitability currently are expected to have higher profitability in the following periods. Empirical evidence from Dechow et al. (1999) and Biddle et al. (2001) is consistent with the premise of our argument.

options (Burgstahler and Dichev 1997; Zhang 2000). More importantly, while growth increases the slope of the value-earnings relation in high-profitability regions, it has either an insignificant or a negative effect on the slope in low-profitability regions.

According to prior studies (Burgstahler and Dichev 1997; Collins et al. 1999), equity value is positively related to equity book value given earnings, but little is known about how growth intervenes in this relation. We explain that, while a positive relation should generally hold absent growth, investment growth will reduce the slope of this relation in high-profitability regions (where growth options are important); in fact, the negative effect of growth can be so dominant that it causes equity value to *decrease* with equity book value. This negative relation in high-profitability regions caused by growth options is contrary to a predicted positive relation in low-profitability regions caused by adaptation options (Burgstahler and Dichev 1997). Together, the two forces produce an equity value function that changes nonmonotonically with book value over a wide profitability range.

Our results support these predictions. Controlling for earnings, we find a negative coefficient on equity book value in the regions of low book value (corresponding to high profitability); however, the slope turns positive in the regions of high book value (low profitability). After explicitly recognizing the growth effect in regressions, we find that growth reduces the slope coefficient (i.e., makes it more negative) on book value in high-profitability regions, but increases the slope (i.e., makes it more positive) in low-profitability regions.

However, the above-described behavior does not portray a full view of the relation between equity value and equity book value in cross sections. If we hold profitability (instead of earnings) constant, then the relation changes sharply. In this case, equity value increases uniformly with book value, and growth increases the slope of this relation. The intuition is that among firms equally efficient in generating value from invested capital, those with more capital invested will have higher values; this is true especially for firms expected to undertake high growth.

Whereas the above effects of investment growth are induced by *economic* forces for value creation, Zhang (2000) explains that investment activities can also influence the value-accounting relation for *accounting* reasons. Under conservative accounting practices, the speed of increased investment in recent years will determine the degree of conservatism in earnings. Consistent with this prediction, we find that the earnings coefficient is greater in the high- versus low-earnings conservatism group (faster investment increases in the recent past). At the same time, we find that the predicted growth effect holds in both high- and low-earnings-conservatism groups, confirming that the economic-driven effect of investment growth is distinct from the conservatism-driven effect of past investments.

Our study extends the empirical valuation literature by showing how the nonlinear relation between equity value, earnings, and equity book value is dependent on investment growth. Previous research has relied heavily on linear valuation relations,[5] motivated by either heuristic arguments or the models of Ohlson (1995) and Feltham and Ohlson (1995), assuming linear information dynamics. However, empirical evidence generally does not support these linear models (Dechow et al. 1999; Callen and Segal 2005). Some criticize these models as nondescriptive because they do not capture positive economic rents or growth options (Lo and Lys 2000; Biddle et al. 2001; Holthausen and Watts 2001). Burgstahler and Dichev (1997) explicitly examine nonlinearity in valuation, but they focus on the role of adaptation options (relevant to unprofitable firms), not that of growth options (relevant mainly to profitable firms).

[5] Examples include Landsman (1986), Barth (1991), and Shevlin (1991), who use balance sheet data, Barth et al. (1992), who use income statement data, and Collins et al. (1999), Lev and Zarowin (1999), and Core et al. (2003), who use a linear combination of earnings and equity book value. See Holthausen and Watts (2001) for a survey.

Our study answers the call by Holthausen and Watts (2001) for more valuation research that incorporates rents and growth options. Although several studies have demonstrated the important role of growth options, their focuses differ from ours. Zhang (2000) builds growth options into a theoretical model, but does not conduct a corresponding empirical examination. Biddle et al. (2001) explore the nonlinear dynamic behavior of residual income under the notion of "capital following profitability" (the key assumption giving rise to real options), but do not examine the behavior of equity value. Finally, Chen and Zhang (2007) test a return model derived from the value model of Zhang (2000). But determining equity return over a time period is a distinctly different problem from determining equity value at a given point in time. Returns derive from changes in value (but one cannot determine a firm's value from observed returns), so the valuation question examined in our study is more basic than that in Chen and Zhang (2007). Importantly, unlike Chen and Zhang (2007), who use the growth variable as a separate explanatory factor—along with their accounting variables—in the return model, our study explores how the valuation impact (the slope coefficients) of accounting variables is dependent on growth, thus emphasizing the interactive effects of growth and accounting variables in valuation.

In addition to examining the role of *prospective* investment (growth options), this study contributes to the literature by showing how *past* investment activities affect the value-accounting relation as a result of conservative accounting practices. Finally, beyond testing the hypotheses of this study, the empirical models we estimate also confirm results established from previous studies concerning, for example, convexity (Burgstahler and Dichev 1997), and the relative importance of earnings versus equity book value for positive and negative earnings firms (similar to our high- and low-profitability firms; Collins et al. 1999). Our study thus helps to unify various empirical findings within a common framework.

In Section II, we develop the hypotheses. Section III describes the sample. Section IV uses partial model specifications to test the hypotheses, while Section V uses full model specifications. Section VI tests the implications of conservative accounting. Finally, we describe our robustness checks in Section VII and conclude in Section VIII. The mathematical derivations of the hypotheses are in the Appendix.

II. THEORETICAL FOUNDATION AND HYPOTHESES

In this section, we first use a simplified version of Zhang's (2000) model, which assumes unbiased accounting depreciation, to develop four hypotheses on the valuation effect of investment growth; these hypotheses all arise from *economic* incentives for value generation. We then consider the general version of Zhang's model to develop our fifth hypothesis, which arises from accounting conservatism (as represented by a conservative depreciation policy in Zhang [2000]).

A Simplified Version of Zhang's (2000) Model

Let V be the (current) equity value of an all-equity-financed firm, B the corresponding equity book value, X the current-period earnings, g the extent to which the scale of operations may grow in the future, and k the earnings capitalization factor.[6] Define $q \equiv X/B$ as the firm's (current-period) profitability. Then, following Zhang (2000) and assuming unbiased accounting depreciation, equity value can be expressed as:

$$V = B[P(q) + kq + gC(q)] = BP(q) + kX + gBC(q) \tag{1}$$

where $P(.)$ is the (put) option to abandon operations, and $C(.)$ is the (call) option to grow, both

[6] Since we are focusing on the role of *contemporaneous* accounting data, we suppress the time subscripts on variables for simplicity.

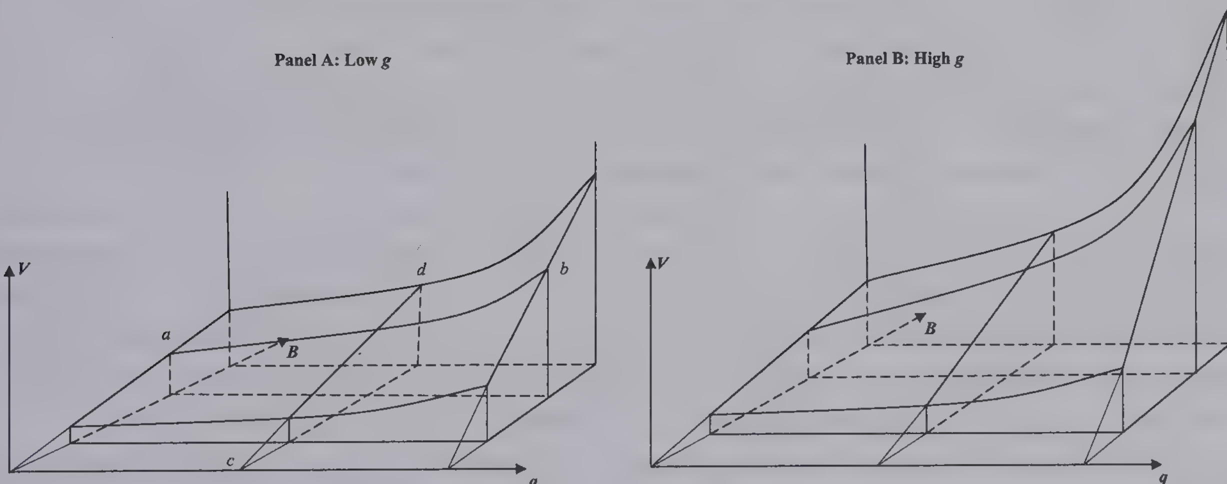
normalized by the scale measure, B. The decisions to exercise the real options are dependent on profitability, making the options functions of q. Panels A and B of Figure 1 provide geometric views of model (1) in the space defined by axes q, B, and V, for two different values of g. The valuation function is represented as a curved surface that maps (q, B) combinations into Vs.[7]

Growth and the Relation between Equity Value and Earnings, Given Book Value

We now employ model (1) to explore how equity value changes with earnings at a given equity book value. The basic V-X relation is depicted in Figure 1, Panel A. At a given B, moving along the q-axis, we trace out V as an increasing and convex function of X (see Predictions 1 and 2 of Zhang [2000, 283]), as represented by curve *ab*. In the extremely low (possibly negative) q region, the slope in the V-X relation is close to zero. The reason is that when operations are so unprofitable that they are likely to be discontinued, equity value depends primarily on book value (proxying for the exit or adaptation value), with earnings being of little use in predicting value generation. As X increases (and hence q increases, given B), the probabilities also increase that the firm will stay in business and its operations will grow. This raises expectations about future earnings, thus increasing the slope in the V-X relation.

In model (1), growth parameter g magnifies the value of the growth option. In high-q regions where the option is in the money, the slope of the V-X relation will increase with growth, as illustrated in Panels A and B of Figure 1 for two different values of g. Moving to the lower-q

FIGURE 1
Geometrical Views of Valuation Model (1)



This figure presents geometrical views of equity value (V) as a function of profitability (q) and book value (B) based on model (1) for low growth (Panel A) and high growth (Panel B). At a given B, V is an increasing function of q, as indicated by curve *ab* in Panel A, and the slope increases with q. At a given q, V is an increasing function of B, as indicated by line *cd* in Panel A.

7 Axes B and q are orthogonal because their information contents are complementary in nature. Note that earnings are dependent on scale, and thus would be correlated with book value.

regions, the growth option becomes less valuable and, consequently, the incremental slope effect of high (versus low) growth diminishes.

In the extreme low-q region, growth should create little or no value. Why firms in this region still undertake growth is outside the model of Zhang (2000), but this can occur for a number of reasons. One is that these firms have temporarily low earnings, such that current profitability is not representative of future profitability. Another is that managers may undertake growth to advance personal interests at the expense of investors. Thus, how growth influences valuation mapping in low-profitability regions is an open empirical issue; to the extent that growth undertaken by these firms does not create value (or destroys value), the growth effect on the slope will disappear (or turn negative).

The above discussion leads to H1 below (stated in alternative form) about the effect of growth as predicted from Zhang (2000). We present qualitative arguments for the hypotheses in this section, and provide the mathematical derivations in the Appendix.

H1: Given equity book value, growth has a positive effect on the slope of the value-earnings relation in high-profitability regions; this effect becomes smaller in lower-profitability regions.

Growth and the Relation between Equity Value and Book Value, Given Earnings

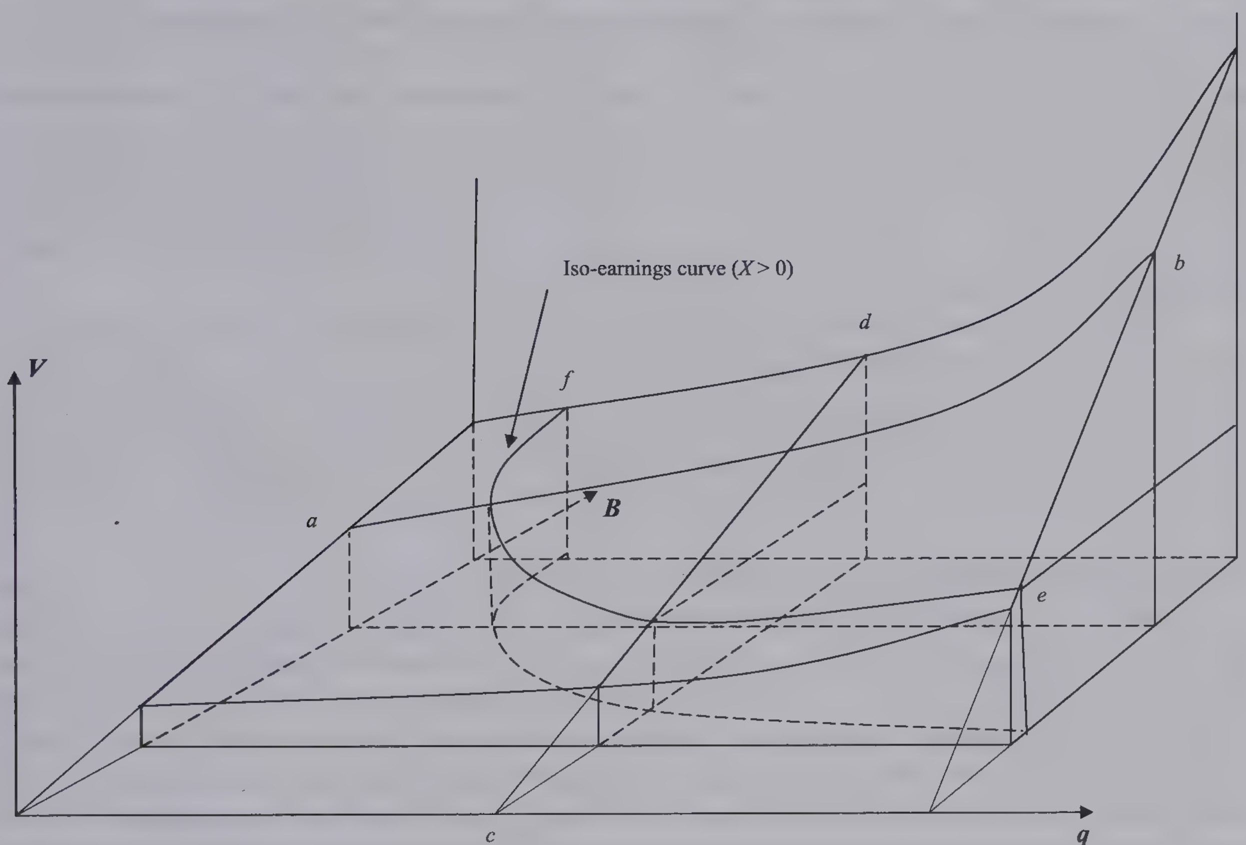
Zhang (2000, 284–285) shows that, given earnings, the relation between equity value and equity book value is nonmonotonic. In the low-q region (where firms have a high B, given X), $P(q)$ is in the money, whereas $C(q)$ is out of the money. Equity value in this case depends primarily on book value, producing a positive relation between V and B. On the other hand, in the high-q region (where firms have a low B), while $P(q)$ is out of the money, $C(q)$ is valuable. Since firms with larger book values are less profitable (holding earnings constant), their growth options are less valuable than those of firms with smaller book values (which are more profitable), causing an inverse relation between V and B in the high-q region. The nonmonotonic relation between V and B, given X, is illustrated by curve *ef* in Figure 2.

Again, growth changes the valuation mapping, with the extent of the growth effect dependent on profitability. As explained above, in the high-q region, the growth option is highly sensitive to profitability, which induces an inverse relation between V and B. Following this reasoning, higher growth will exacerbate the effect of growth options, causing the inverse V-B relation more pronounced. This suggests that growth should have a negative effect on the slope of the V-B relation (i.e., make the slope more negative) in the high-q region.

The situation is different in the low-q region. The scenario of low-profitability firms undertaking growth lies outside Zhang's (2000) model, and how growth influences valuation mapping is an empirical issue. Intuitively, investment growth increases the proportion of newly acquired assets relative to the old assets of a firm. The relevant question then is: In the event of abandonment, how easily are new assets adapted to alternative business uses relative to old assets? To the extent that the former are more adaptable and, thus, have higher exit values than the latter (on a per-unit basis), a higher proportion of new assets will increase the percentage of asset values to be recovered, suggesting that growth will increase the slope of the V-B relation in the low-q region.

The above discussion leads to H2 and H3 below. Note that H2 and part (i) of H3 are based on Zhang (2000), whereas part (ii) of H3 is a prediction outside Zhang's (2000) model.

FIGURE 2
Relation between Equity Value and Equity Book Value, Given Earnings



This figure illustrates that equity value (V) is a nonmonotonic function of book value (B), given earnings (X). Curve *ef* is an iso-earnings curve. As we move along curve *ef*, starting from point *e*, V first decreases with B in the high-profitability (low-book-value) region and then increases with B in the low-profitability (high-book-value) region.

H2: Given earnings, equity value increases with equity book value in low-profitability (high-book-value) regions, and decreases with book value in high-profitability (low-book-value) regions where growth options are important.[8]

H3: (i) Given earnings, growth reduces the slope of the relation between equity value and book value in high-profitability regions (i.e., makes the slope more negative); and (ii) growth increases this slope in low-profitability regions (i.e., makes the slope more positive).

Growth and the Relation between Equity Value and Book Value, Given Profitability

The relation between equity value and book value can change sharply, however, if we hold profitability constant, rather than earnings. Refer to Figure 1, Panel A. For a given q, moving along

[8] This nonmonotonic V-B relation, given X, is in contrast to the prediction of Burgstahler and Dichev (1997, their Proposition 2) that equity value is an increasing function of book value given earnings, arising from a setting with adaptation options but no growth options.

axis B we trace out V as an increasing function of B, as represented by the line *cd*. This means that, among firms equally capable of generating profit from invested capital, those with more invested capital have higher values, resulting in a positive *V-B* relation given q. This behavior of the *V-B* relation is distinct from that explained above, where we control for X in the cross-section.

The influence of growth on the *V-B* relation is also different here. *Ceteris paribus*, the operations of firms with more growth opportunities will grow faster than will those of firms with fewer opportunities, which will, in turn, generate more earnings. Thus, growth increases the slope in the *V-B* relation, given q, in the high-q region. In the low-q region, a more relevant notion of value is adaptation value. On the premise that newly invested assets are more adaptable to alternative business uses than are older assets (as discussed above), we conjecture that growth also increases the slope of the *V-B* relation in the low-q region.

We summarize this discussion in H4. Note that part (i) is based on Zhang (2000), whereas part (ii) is an intuitive prediction outside Zhang's (2000) model.

H4: (i) In high-profitability regions, growth increases the (positive) slope in the relation between equity value and equity book value, given profitability; and (ii) in low-profitability regions, growth also increases the slope in this relation.

Past Investment Activities and Conservatism in Earnings

Whereas the above analysis illustrates the valuation effect of *prospective* investment activities that arises from economic incentives for value creation, the general version of Zhang's model (Zhang 2000, 279, Equation 24) further incorporates conservative accounting, giving rise to the role of *past* investments in influencing the valuation function. In Zhang (2000), accounting is based on historical cost, and the source of accounting biases is a conservative policy that recognizes depreciation at a rate *faster* than the actual decline of the productive capacity of assets. Unbiased accounting obtains in the special case in which depreciation is recognized in an unbiased way, commensurate with the actual decline in physical asset stock.[9]

Specifically, under a conservative depreciation policy, firms recognize more expense (relative to unbiased depreciation) in the initial years of an asset's life, but less expense in the latter years when the cumulative over-recognition gets unwound. In a given year, the direction and magnitude of earnings bias depend jointly on the amount of depreciation over-recognition on newly invested assets and the amount of under-recognition on older assets (i.e., the unwinding of previous over-recognition). Following periods of *increases* in capital investments, the depreciation over-recognition on newly invested assets (which are in greater amounts) outweighs the depreciation under-recognition on older assets (which are in smaller amounts), causing the firm's overall earnings to be understated (relative to the case of unbiased depreciation). On the other hand, following periods of *decreases* in investments, the opposite result occurs, whereby the depreciation over-recognition on newly invested assets is outweighed by the unwinding of previous over-recognition on older assets, causing the firm's overall earnings to be overstated.[10]

---

9 The notion of conservatism here is similar to that in Penman and Zhang (2002) and Rajan et al. (2007). With historical cost-based accounting, equity book value relates to realized investments but not to expected future activities. In such a setting, even under unbiased accounting, equity book value on average will be below equity market value owing to (for example) expected positive NPV investments in the future. An alternative and fundamentally different notion of conservatism is proposed by Ohlson (1995) and Feltham and Ohlson (1995), which is in terms of whether equity book value is expected to be below or equal to equity market value. Under this latter notion, unbiased accounting requires that equity book value be marked to market, and consequently income be determined as changes in equity market value; this requires that equity book value incorporate not only past investments, but also the expected value from future investments.

10 See Zhang (2000, 278–279) for a more detailed analysis. The link between past investment and earnings biases is similarly illustrated by Rajan et al. (2007).

However, investors will correct for accounting biases in valuation. Following the above discussion, we predict an increased valuation weight on earnings following periods of increases in investments (because earnings are understated) and a reduced weight following periods of decreases in investments (because earnings are overstated) (see Zhang 2000, 283, Hypothesis 1). This leads to our fifth hypothesis, which concerns the relation between the trend of past investment changes and the slope of the value-earnings relation.[11]

H5: *Ceteris paribus*, the slope of the value-earnings relation is greater for firms experiencing fast investment increases in recent periods than for firms experiencing slow (or negative) investment increases.

III. SAMPLE AND DESCRIPTIVE STATISTICS

We extract the data for empirical analysis from the Compustat database. The variable definitions and specific data items are as follows: V (market value per share) is the market price of common shares (Compustat data item #199) at the fiscal year-end; B (book value per share) is the book value of equity (#60) divided by the number of common shares outstanding (#25), both at the fiscal year-end; X (earnings per share) is diluted earnings per share excluding extraordinary items (#57);[12] and q (profitability) is earnings before extraordinary items divided by the book value of equity at the beginning of the fiscal year.[13] Growth in Zhang's (2000) model refers to prospective investment growth. Empirically, we measure growth (g) by the average realized annual growth rate of equity book value, adjusted for the Consumer Price Index, over the subsequent three years (similar to the growth measure in Biddle et al. [2001]).[14]

We require firms to have at least four consecutive years of data to be included in the sample (due to using *ex post* growth measures). We delete observations with missing data on earnings per share (X), book value per share (B), market value per share (V), profitability (q), or growth (g). We also remove regulated and financial industries (as in Burgstahler and Dichev 1997). To mitigate the effect of outliers, we exclude firms with a current- or prior-year total equity book value of less than $1 million, and winsorize all continuous variables in our regressions at the top and bottom 1 percent of the distributions. These steps result in a sample of 101,672 observations covering the period 1966–2003.

Table 1, Panel A reports the distributional statistics of the overall sample. Market value per share (V) ranges from 0.094 to 143.125, with a mean (median) of 17.718 (12.250), all in dollars; book value per share (B) from 0.034 to 86.308, with a mean (median) of 10.918 (7.601); earnings per share (X) from −7.614 to 10.602, with a mean (median) of 1.055 (0.758); and profitability (q) (i.e., return on beginning book equity) from −3.083 to 1.166, with a mean (median) of 0.069 (0.112). The growth rate in equity book value (g) ranges from a low of −0.515 to a high of 4.004, with a mean (median) of 0.115 (0.041).

Table 1, Panel B provides the pairwise correlations among the variables. The market value of equity is significantly correlated with equity book value (with a correlation of 0.713), earnings (0.703), profitability (0.463), and growth (0.072), thus providing a preliminary indication of the valuation relevance of these variables. We also note that book value is highly correlated with

---

[11] In Zhang (2000), equity book value is always understated under conservative accounting practices, and no cross-sectional relation is predicted between past investments and the valuation impact of equity book value.

[12] Our results are qualitatively unchanged if we alternatively use undiluted earnings per share, defined as earnings before extraordinary items (#18) divided by the number of common shares outstanding at the fiscal year-end (#25).

[13] Our results remain similar if we use year-end book values to define profitability.

[14] Compared with a measure of (output-based) earnings growth, our (input-based) growth proxy matches more closely with the construct in the original theoretical model in Zhang (2000). In robustness checks, we also use an *ex ante* growth measure and measures without inflation adjustment, and find the results similar.

TABLE 1

Descriptive Statistics

Panel A: Distributional Statistics

Variable	Mean	Median	Std.	Min.	Q1	Q3	Max.
V	17.718	12.250	17.691	0.094	5.000	24.750	143.125
B	10.918	7.601	10.894	0.034	3.509	14.507	86.308
X	1.055	0.758	1.717	−7.614	0.057	1.777	10.602
q	0.069	0.112	0.289	−3.083	0.019	0.190	1.166
g	0.115	0.041	0.366	−0.515	−0.036	0.154	4.004

Panel B: Correlations

Variable	V	B	X	q
B	0.713[a]			
X	0.703[a]	0.713[a]		
q	0.463[a]	0.238[a]	0.723[a]	
g	0.072[a]	−0.192[a]	0.022[a]	0.195[a]

[a] Indicates significance at the 0.01 level.

The sample consists of 101,672 firm-year observations from Compustat for the period 1966–2003. Market value per share (V) is the market price of common shares (Compustat data item #199). Book value per share (B) is calculated as the total book value of equity (#60) divided by the number of shares outstanding (#25) at the fiscal year-end. Earnings per share (X) is (diluted) earnings per share excluding extraordinary items (#57).

Profitability (q) is income before extraordinary items (#18) divided by the beginning book value of equity.

Growth opportunity (g) is the average annual growth rate of equity book value over the subsequent three years.

earnings, with a correlation of 0.713, versus 0.238 with profitability. The growth rate is significantly correlated with profitability, with a correlation of 0.195, confirming the economic intuition of "capital following profitability" (Biddle et al. 2001). Growth has a negative correlation with book value (−0.192).

IV. RESULTS BASED ON PARTIAL MODEL SPECIFICATIONS

To examine the valuation effects of investment growth (H1 to H4), we perform two sets of tests. One set of tests, provided in this section, uses "partial" specifications of the underlying valuation model, whereby we explicitly control for the variable needing to be held constant in a hypothesis. Partial models collapse one of the dimensions of the valuation space to view the valuation model from a particular angle. Thus, these simplified models highlight particular aspects of the growth effect. The limitation is that the individual hypotheses (concerning different aspects of the growth effect) are tested in separate regressions. The second set of tests, reported in subsequent sections, uses full model specifications, which allow multiple hypotheses to be tested jointly. This latter approach requires more complex specifications to integrate different parts of the growth effect into one regression.

When using partial model specifications, different methods are available to keep the control variable constant. Here, we choose to normalize (scale) the regression variables by the control variable, as in Burgstahler and Dichev (1997).[15] Because scaling by a negative number is not meaningful, for the analyses in this section we limit the sample to those observations with positive earnings (the sample size becomes 79,812).

[15] In robustness checks, we also control for a variable by partitioning samples on the control variable into deciles and run separate regressions for the deciles, finding similar results.

To mitigate multicollinearity concerns, where applicable, we demean the continuous variables in the interaction terms in regressions in this section and in subsequent sections (Aiken and West 1991).

Results on the Relation between Equity Value and Earnings Given Book Value (H1)

H1 concerns the relation between V and X, given B. We first run the following regression each year from 1966 to 2003 to examine the basic V-X relation (which serves as a base model for introducing the growth variable later):

$$V_i/B_i = L_q * (\alpha_0 + \alpha_1 X_i/B_i) + M_q * (\beta_0 + \beta_1 X_i/B_i) + H_q * (\gamma_0 + \gamma_1 X_i/B_i) + e_i, \tag{2}$$

where V_i / B_i is the (per share) stock price of firm i at the end of a fiscal year scaled by the corresponding equity book value, and X_i / B_i is the earnings per share of firm i of a year scaled by the same equity book value. We use indicator variables L_q, M_q, and H_q to indicate observations in the high, medium, and low thirds of profitability in an annual sample, respectively. This specification follows Sirri and Tufano (1998), taking into account cross-correlations among the partitions. Following the predictions of an increasing and convex relation between V and X given B, we expect $0 \leq \alpha_1 \leq \beta_1 \leq \gamma_1$ with at least one of the inequalities strict.

Table 2 reports the average coefficients in Regression (2) across the 38 sample years, along

TABLE 2

Relation between Equity Value and Earnings, Given Equity Book Value

q-partitions	Median *q*	Intercept	*X/B*	Predicted Sign
L_q	0.061	1.302[a]	2.322[a]	+/0
		(9.58)	(4.79)	
M_q	0.142	0.477[a]	11.607[a]	+
		(3.12)	(19.66)	
H_q	0.255	0.086	15.424[a]	+
		(0.26)	(10.69)	
Diff. in *X/B* Coeff.: $M_q - L_q$			9.285[a]	+
			(10.48)	
Diff. in *X/B* Coeff.: $H_q - M_q$			3.817[a]	+
			(3.95)	
Diff. in *X/B* Coeff.: $H_q - L_q$			13.102[a]	+
			(9.23)	
Adj. R^2		0.72		

[a] Indicates one-tailed significance at the 0.01 level.

This table reports the results of Regression (2):

$$V_i/B_i = L_q * (\alpha_0 + \alpha_1 X_i/B_i) + M_q * (\beta_0 + \beta_1 X_i/B_i) + H_q * (\gamma_0 + \gamma_1 X_i/B_i) + e_i.$$

We partition annual samples of positive earnings firms by profitability (q) into thirds and run regressions year by year, from 1966 to 2003. L_q (M_q, H_q) is an indicator variable for observations in the low- (medium-, high-) profitability region. The reported results are the average coefficients across the years, and t-values (in parentheses) are based on Fama-MacBeth standard errors with Newey-West adjustments.

with the Fama-MacBeth t-values with Newey-West adjustments. The slope coefficient on earnings is significantly positive in all profitability regions of our sample (restricted to positive-earnings observations in this section). The coefficient value is 2.322 in the low-profitability region, and increases to 11.607 and 15.4224 in the medium- and high-profitability regions, respectively. The difference in the slope coefficient between the high- and the low-profitability region (13.102) is significant (t = 9.23; $p < 0.01$). This initial analysis confirms the prior findings of an increasing and convex relation between V and X, controlling for B (e.g., Burgstahler and Dichev 1997; Collins et al. 1999).

To test H1, we extend Regression (2) to distinguish between low- and high-growth firms, as follows:

$$V_i/B_i = L_q * (\alpha_0 + \alpha_1 H_g + \alpha_2 X_i/B_i + \alpha_3 H_g X_i/B_i) + M_q * (\beta_0 + \beta_1 H_g + \beta_2 X_i/B_i + \beta_3 H_g X_i/B_i) + H_q * (\gamma_0 + \gamma_1 H_g + \gamma_2 X_i/B_i + \gamma_3 H_g X_i/B_i) + e_i, \quad (3)$$

where H_g is an indicator variable for observations in the high- (versus low-) growth group. Following H1, we expect $\gamma_3 > 0$, and $\gamma_3 \geq \beta_3 \geq \alpha_3$, with at least one of the inequalities strict.

Table 3 provides the results of Regression (3). In the high-profitability region, the incremental effect of high (versus low) growth on the earnings coefficient, captured by γ_3, has a positive value of 4.036 (t = 5.44), significant at the 0.01 level,[16] indicating that, in this region, growth significantly increases the slope of the value-earnings relation, controlling for equity book value. The effect of high (versus low) growth in the medium-profitability region decreases to 1.263 (t = 2.59), but remains significant at the 0.01 level. On the other hand, the effect of high growth in the low-profitability region becomes negative (−2.936), significant at the 0.01 level (t = −8.32). The difference in the slope effect of growth between the high- and medium-profitability regions is significant at the 0.05 level (t = 2.41), and that between the medium- and the low-profitability regions is significant at the 0.01 level (t = 5.39).

These results indicate that investment growth generally has a significant impact on the relation between equity value and earnings, given book value, but that the extent of the growth impact is contingent on the profitability level. Growth increases the value impact of earnings in high-profitability regions, but the effect becomes smaller and turns negative in lower-profitability regions, consistent with H1.

Results on the Relation between Equity Value and Book Value Given Earnings (H2 and H3)

H2 and H3 concern the relation between V and B, holding X constant. We first test H2, where the effect of investment growth is implied but not made explicit in the predicted equity value behavior. Following a design similar to that described above, we scale V and B by X and run the following regression each year from 1966 to 2003:

$$V_i/X_i = L_q * (\alpha_0 + \alpha_1 B_i/X_i) + M_q * (\beta_0 + \beta_1 B_i/X_i) + H_q * (\gamma_0 + \gamma_1 B_i/X_i) + e_i. \quad (4)$$

According to H2, the slope coefficient on book value will be positive in the low-q region, $\alpha_1 > 0$, and negative in the high-q region, $\gamma_1 < 0$. H2 also predicts $\alpha_1 \geq \beta_1 \geq \gamma_1$, with at least one of the inequalities strict.

Table 4 reports the results of Regression (4). As predicted, the coefficient on equity book value has a positive value of 0.668 (t = 8.05) in the low-profitability region, significant at the 0.01 level, whereas it has a negative value of −0.432 (t = −2.39) in the high-profitability region, significant at the 0.05 level. The coefficient in the medium region is 0.386 (t = 2.39), which is

[16] Throughout the paper, we determine statistical significance with Bonferroni corrections, which require the hypothesized conditions to hold jointly in a regression.

TABLE 3

Growth and the Relation between Equity Value and Earnings, Given Equity Book Value (H1)

q-partitions	Median q	Median g: $H_g=0$	Median g: $H_g=1$	Intercept	H_g	X/B	H_gX/B	Predicted Sign	$X/B+H_gX/B$
L_q	0.061	−0.036	0.133	1.219[a]	0.601[a]	2.593[a]	−2.936[a],***	+/−	−0.343
				(10.54)	(6.60)	(5.98)	(−8.32)		(−0.36)
					[1.01]	[1.49]	[1.49]		
M_q	0.142	−0.014	0.116	1.624[a]	0.450[a]	9.557[a]	1.263[a],***	+	10.820[a]
				(10.81)	(8.39)	(19.21)	(2.59)		(10.36)
					[1.01]	[2.01]	[2.00]		
H_q	0.255	−0.018	0.165	2.808[a]	0.755[a]	11.853[a]	4.036[a],***	+	15.889[a]
				(9.83)	(12.56)	(13.82)	(5.44)		(13.12)
					[1.04]	[4.08]	[4.02]		
Diff. in *X/B* Coeff.: $M_q - L_q$							4.199[a],***	+	
							(5.39)		
Diff. in *X/B* Coeff.: $H_q - M_q$							2.773[a],**	+	
							(2.41)		
Diff. in *X/B* Coeff.: $H_q - L_q$							6.972[a],***	+	
							(9.55)		
Adj. R^2						0.73			

[a] Indicates one-tailed significance at the 0.01 level without Bonferroni corrections.

, * Indicate one-tailed significance at the 0.05 and 0.01 levels with Bonferroni corrections (m, the number of parts in the hypothesis, is 3), respectively.

This table reports the results of Regression (3):

$$V_i/B_i = L_q * (\alpha_0 + \alpha_1 H_g + \alpha_2 X_i/B_i + \alpha_3 H_g X_i/B_i) + M_q * (\beta_0 + \beta_1 H_g + \beta_2 X_i/B_i + \beta_3 H_g X_i/B_i) + H_q * (\gamma_0 + \gamma_1 H_g + \gamma_2 X_i/B_i + \gamma_3 H_g X_i/B_i) + e_i$$

where H_g is an indicator variable equal to 1 for firms in the high half of growth (measured by the average annual growth in equity book value in the subsequent three years) in a sample. The explanatory factors that are continuous variables are mean-adjusted. We partition annual samples of positive earnings firms by profitability (q) into thirds and run regressions year by year, from 1966 to 2003. L_q (M_q, H_q) is an indicator variable for observations in the low- (medium-, high-) profitability region. The reported results are the average coefficients across the years, and t-values (in parentheses) are based on Fama-MacBeth standard errors with Newey-West adjustments. The numbers in brackets are Variance Inflation Factors (VIF).

between those in the two extreme regions. The difference in the slope effect of growth between the high- and the medium-profitability regions is significant at the 0.01 level (t = −13.66), and that between the medium- and the low-profitability regions is also significant at the 0.01 level (t = −3.46). These results are consistent with the prediction in H2 that, given earnings, equity value is inversely related to equity book value among firms with high profitability, and positively related to book value among firms with low profitability.

To test H3, we extend Regression (4) to explicitly build in the growth effect in Regression (5) below:

TABLE 4

Relation between Equity Value and Equity Book Value, Given Earnings (H2)

q-partitions	Median *q*	Intercept	*B/X*	Predicted Sign
L_q	0.061	9.086[a]	0.668[a],***	+
		(8.60)	(8.05)	
M_q	0.142	11.505[a]	0.386[a],**	+/−
		(30.05)	(2.39)	
H_q	0.255	17.715[a]	−0.432[a],**	−
		(21.64)	(−2.39)	
Diff. in *B/X* Coeff.: $M_q - L_q$			−0.282[a],***	−
			(−3.46)	
Diff. in *B/X* Coeff.: $H_q - M_q$			−0.817[a],***	−
			(−13.66)	
Diff. in *B/X* Coeff.: $H_q - L_q$			−1.100[a],***	−
			(−10.25)	
Adj. R^2			0.78	

[a] Indicates one-tailed significance at the 0.01 level without Bonferroni corrections.

, * Indicate one-tailed significance at the 0.05 and 0.01 levels with Bonferroni corrections (m = 4), respectively.

This table reports the results of Regression (4):

$$V_i/X_i = L_q * (\alpha_0 + \alpha_1 B_i/X_i) + M_q * (\beta_0 + \beta_1 B_i/X_i) + H_q * (\gamma_0 + \gamma_1 B_i/X_i) + e_i.$$

We partition annual samples of positive earnings firms by profitability (q) into thirds and run regressions year by year, from 1966 to 2003. L_q (M_q, H_q) is an indicator variable for observations in the low- (medium-, high-) profitability region. The reported results are the average coefficients across the years, and t-values (in parentheses) are based on Fama-MacBeth standard errors with Newey-West adjustments.

$$V_i/X_i = L_q * (\alpha_0 + \alpha_1 H_g + \alpha_2 B_i/X_i + \alpha_3 H_g B_i/X_i) + M_q * (\beta_0 + \beta_1 H_g + \beta_2 B_i/X_i + \beta_3 H_g B_i/X_i)$$
$$+ H_q * (\gamma_0 + \gamma_1 H_g + \gamma_2 B_i/X_i + \gamma_3 H_g B_i/X_i) + e_i. \quad (5)$$

According to H3, growth will increase the book value coefficient (i.e., make it more positive) when profitability is low ($\alpha_3 > 0$), but decrease it (i.e., make it more negative) when profitability is high ($\gamma_3 < 0$). H3 also predicts $\alpha_3 \geq \beta_3 \geq \gamma_3$ with at least one of the inequalities strict.

Table 5 reports the results of Regression (5). As predicted, in the low-profitability region, the effect of high growth on the book value coefficient is positive (0.163) and significant at the 0.01 level (t = 3.97), while its effect in the high-profitability region is negative (−0.374) and significant at the 0.01 level (t = −2.82). In the medium-profitability region, the effect of high growth is positive (0.114), but not statistically significant (t = 0.91). The difference of −0.537 in the slope effect between the high- and the low-profitability region is significant at the 0.01 level (t = −3.50). These results are consistent with H3, which predicts that growth will exhibit opposite effects on the *V-B* relation given *X* in the high- and low-profitability regions.

Results on the Relation between Equity Value and Book Value Given Profitability (H4)

According to H4, the *V-B* relation behaves very differently if we hold *q*, rather than *X*, constant. We adopt Equation (6) to examine the basic relation between *V* and *B*, controlling for *q*:

TABLE 5

Growth and the Relation between Equity Value and Equity Book Value, Given Earnings (H3)

q-partitions	Median q	Median g: $H_g=0$	Median g: $H_g=1$	Intercept	H_g	B/X	H_gB/X	Predicted Sign	$B/X+H_gB/X$
L_q	0.061	−0.036	0.133	20.165[a]	4.893[a]	0.654[a]	0.163[a],***	+	0.817[a]
				(7.30)	(6.15)	(7.88)	(3.97)		(11.93)
					[1.02]	[1.42]	[1.43]		
M_q	0.142	−0.014	0.116	13.592[a]	2.580[a]	0.434[a]	0.114	+/−	0.548[a]
				(11.02)	(7.85)	(4.29)	(0.91)		(3.73)
					[1.01]	[1.98]	[1.97]		
H_q	0.255	−0.018	0.165	13.281[a]	2.767[a]	−0.035	−0.374[a],***	−	−0.409[a]
				(13.11)	(11.65)	(−0.47)	(−2.82)		(−3.77)
					[1.02]	[3.23]	[3.20]		
Diff. in H_g B/X Coeff.: M_q-L_q							−0.049	−	
							(−0.43)		
Diff. in H_g B/X Coeff.: H_q-M_q							−0.488[a],***	−	
							(−2.74)		
Diff. in H_g B/X Coeff.: H_q-L_q							−0.537[a],***	−	
							(−3.50)		
Adj. R^2							0.79		

[a] Indicates one-tailed significance at the 0.01 level without Bonferroni corrections.

*** Indicates one-tailed significance at the 0.01 level with Bonferroni corrections (m = 4).

This table reports the results of Regression (5):

$$V_i/X_i = L_q*(\alpha_0+\alpha_1H_g+\alpha_2B_i/X_i+\alpha_3H_gB_i/X_i)+M_q*(\beta_0+\beta_1H_g+\beta_2B_i/X_i+\beta_3H_gB_i/X_i)+H_q*(\gamma_0+\gamma_1H_g+\gamma_2B_i/X_i+\gamma_3H_gB_i/X_i)+e_i,$$

where H_g is an indicator variable equal to 1 for firms in the high half of growth (measured by the average annual growth in equity book value in the subsequent three years) in a sample. The explanatory factors that are continuous variables are mean-adjusted. We partition annual samples of positive earnings firms by profitability (q) into thirds and run regressions year by year, from 1966 to 2003. L_q (M_q, H_q) is an indicator variable for observations in the low- (medium-, high-) profitability region. The reported results are the average coefficients across the years, and t-values (in parentheses) are based on Fama-MacBeth standard errors with Newey-West adjustments. The numbers in brackets are Variance Inflation Factors (VIF).

$$V_i/q_i = L_q*(\alpha_0 + \alpha_1 H_b + \alpha_2 B_i/q_i + \alpha_3 H_b B_i/q_i) + M_q*(\beta_0 + \beta_1 H_b + \beta_2 B_i/q_i + \beta_3 H_b B_i/q_i) + H_q*(\gamma_0 + \gamma_1 H_b + \gamma_2 B_i/q_i + \gamma_3 H_b B_i/q_i) + e_i, \quad (6)$$

where H_b is an indicator variable for observations with equity book value above the sample median, which allows for possible changes in the slope across different B-regions.

Table 6 reports the results of Regression (6). The coefficient on equity book value is significantly positive in all three q-partitions. For low-q firms, the slope coefficient is 0.898 (t = 9.50) in the low-book-value region and 0.823 (t = 19.65) in the high-book-value region. These coefficients are 1.187 (t = 8.77) and 0.702 (t = 20.10) for medium-q firms, and 1.968 (t = 9.61) and 0.921 (t = 12.26) for high-q firms. These results confirm the basic intuition that, holding profitability constant, equity value will be higher for firms with more capital invested in operations. We also find that the slope becomes smaller as book value increases, mainly in higher profitability partitions. This may be a sign of diminishing returns to scale, indicating that the value derived from additional investment declines as scale expands.

To test H4, we modify Regression (6) to incorporate the growth effect as follows:

$$V_i/q_i = L_q*(\alpha_0 + \alpha_1 H_g + \alpha_2 B_i/q_i + \alpha_3 H_g B_i/q_i) + M_q*(\beta_0 + \beta_1 H_g + \beta_2 B_i/q_i + \beta_3 H_g B_i/q_i) + H_q*(\gamma_0 + \gamma_1 H_g + \gamma_2 B_i/q_i + \gamma_3 H_g B_i/q_i) + e_i. \quad (7)$$

As per H4, the incremental slope coefficient on book value is positive in all profitability ranges, $\alpha_3 > 0$, $\beta_3 > 0$, and $\gamma_3 > 0$.

Table 7 reports the results. The base coefficient on book value is significantly positive, equal to 0.884 (t = 9.20), 1.078 (t = 11.86), and 1.508 (t = 12.23) in the low-q, medium-q, and high-q regions, respectively. More importantly, the incremental slope effect of high growth is also significantly positive, equal to 0.050 (t = 5.19), 0.139 (t = 9.38), and 0.291 (t = 9.85), respectively, in these profitability regions, which further increases the coefficient on equity book value. These results are consistent with the prediction in H4 that growth will increase the (positive) slope in the relation between equity value and book value, given profitability.

A comparison between the results in Tables 6 and 7 here and those in Tables 4 and 5 above demonstrates that the behavior of the V-B relation is highly sensitive to whether we control for X or q in cross sections.

V. RESULTS BASED ON FULL MODEL SPECIFICATIONS

Thus far, we have tested H1 through H4 individually using partial models. We now test the hypotheses using "full" model specifications in which X and B may vary simultaneously. The use of a full model specification enables a more complete view of the behavior of equity value, allowing for testing the different aspects of the growth effect jointly within a common equation.

To design a regression specification for this purpose, we first rewrite theoretical model (1) as $V/B = P(q) + kq + gC(q)$, which is increasing and convex in q, and then proxy it empirically by a piecewise linear function of q that further distinguishes between high- and low-growth firms using H_g. Next, we multiply both sides of the simplified equation by B and rearrange the terms to recognize $qB = X$. These steps yield the following regression for explaining equity value:

$$V_i = \alpha_0 + \alpha_1 H_g + \alpha_2 H_q + \alpha_3 H_g H_q + \beta_0 B_i + \beta_1 H_g B_i + \beta_2 H_q B_i + \beta_3 H_g H_q B_i + \gamma_0 X_i + \gamma_1 H_g X_i + \gamma_2 H_q X_i + \gamma_3 H_g H_q X_i + e_i. \quad (8)$$

We run Regression (8) on annual samples, including firms with negative earnings as well as those with positive earnings, and calculate the average coefficients across the sample years together with the Fama-MacBeth t-values with Newey-West adjustments. As before, to mitigate multicollinear-

TABLE 6

Relation between Equity Value and Book Value, Given Profitability

q-partitions	Median q	Median B: $H_b = 0$	Median B: $H_b = 1$	Intercept	H_b	B/q	Predicted Sign	H_bB/q	$B/q + H_bB/q$	Predicted Sign
L_q	0.061	4.603	17.611	284.399[a]	−10.639[a]	0.898[a]	+	−0.075[a]	0.823[a],***	+
				(15.80)	(−3.93)	(9.50)		(−5.81)	(19.65)	
					[1.47]	[1.47]		[1.04]		
M_q	0.142	5.421	17.177	178.769[a]	1.591	1.187[a]	+	−0.485[a]	0.702[a],***	+
				(20.98)	(0.31)	(8.77)		(−4.70)	(20.10)	
					[9.30]	[9.68]		[7.40]		
H_q	0.255	4.871	14.477	117.600[a]	8.665[a]	1.968[a]	+	−1.047[a]	0.921[a],***	+
				(16.92)	(8.21)	(9.61)		(−4.19)	(12.26)	
					[2.85]	[8.83]		[5.70]		
Adj. R^2							0.87			

[a] Indicates one-tailed significance at the 0.01 level.

*** Indicates one-tailed significance at the 0.01 level with Bonferroni corrections (m = 3).

This table reports the results of Regression (6):

$$V_i/q_i = L_q * (\alpha_0 + \alpha_1 H_b + \alpha_2 B_i/q_i + \alpha_3 H_b B_i/q_i) + M_q * (\beta_0 + \beta_1 H_b + \beta_2 B_i/q_i + \beta_3 H_b B_i/q_i) + H_q * (\gamma_0 + \gamma_1 H_b + \gamma_2 B_i/q_i + \gamma_3 H_b B_i/q_i) + e_i$$

where H_b is an indicator variable equal to 1 for firms in the high half of book value in a sample. The explanatory factors that are continuous variables are mean-adjusted. We partition annual samples of positive earnings firms by profitability (q) into thirds and run regressions year by year, from 1966 to 2003. L_q (M_q, H_q) is an indicator variable for observations in the low- (medium-, high-) profitability region. The reported results are the average coefficients across the years, and t-values (in parentheses) are based on Fama-MacBeth standard errors with Newey-West adjustments. The numbers in brackets are Variance Inflation Factors (VIF).

TABLE 7

Growth and the Relation between Equity Value and Equity Book Value, Given Profitability (H4)

q-partitions	Median q	Median g: $H_g=0$	Median g: $H_g=1$	Intercept	H_g	B/q	H_gB/q	Predicted Sign	$B/q+H_gB/q$
L_q	0.061	−0.036	0.133	267.631[a]	40.823[a]	0.884[a],***	0.050[a],***	+	0.934[a],***
				(16.76)	(7.74)	(9.20)	(5.19)		(20.69)
					[1.07]	[1.44]	[1.48]		
M_q	0.142	−0.014	0.116	158.428[a]	19.438[a]	1.078[a],***	0.139[a],***	+	1.217[a],***
				(19.94)	(9.07)	(11.86)	(9.38)		(21.66)
					[1.03]	[1.72]	[1.71]		
H_q	0.255	−0.018	0.165	93.935[a]	17.012[a]	1.508[a],***	0.291[a],***	+	1.799[a],***
				(18.13)	(4.15)	(12.23)	(9.85)		(22.51)
					[1.03]	[2.61]	[2.57]		
Adj. R^2						0.88			

[a] Indicates one-tailed significance at the 0.01 level.

*** Indicates one-tailed significance at the 0.01 level with Bonferroni corrections (m = 6).

This table reports the results based on Regression (7):

$$V_i/q_i = L_q * (\alpha_0 + \alpha_1 H_g + \alpha_2 B_i/q_i + \alpha_3 H_g B_i/q_i) + M_q * (\beta_0 + \beta_1 H_g + \beta_2 B_i/q_i + \beta_3 H_g B_i/q_i) + H_q * (\gamma_0 + \gamma_1 H_g + \gamma_2 B_i/q_i + \gamma_3 H_g B_i/q_i) + e_i,$$

where the explanatory factors that are continuous variables are mean-adjusted. We partition annual samples of positive earnings firms by profitability (q) into thirds and run regressions year by year, from 1966 to 2003. L_q (M_q, H_q) is an indicator variable for observations in the low- (medium-, high-) profitability region. The reported results are the average coefficients across the years, and t-values (in parentheses) are based on Fama-MacBeth standard errors with Newey-West adjustments. The numbers in brackets are Variance Inflation Factors (VIF).

ity concerns, we use demeaned X and B in the regressions. The results of Regression (8) are reported in Table 8.

Testing H1, H2, and H3

Regression (8) allows for directly testing H1, H2, and H3, but not H4 (explained below). Our analysis here concerns whether the different parts of the growth effect stated in H1, H2, and H3 hold jointly, which amounts to a test of the "total" growth effect as represented by these hypotheses.[17] A joint test of these hypotheses requires stricter conditions for accepting them than in separate tests (through a Bonferroni adjustment), so accepting the hypotheses jointly automatically implies accepting them separately.

H1 predicts that growth will increase the slope of the *V-X* relation, given B, in the high-profitability region, $\gamma_1 + \gamma_3 > 0$, and that the growth effect on this slope will be greater in the high- than in the low-profitability region, $\gamma_3 > 0$. In Table 8, the estimate of $\gamma_1 + \gamma_3$ is 1.811 (t = 2.31), and that of γ_3 is 1.860 (t = 3.46). These coefficients have the predicted signs and are significant at the 0.1 level or better.

H2 predicts that (i) given X, V will decrease with B in the high-profitability region, which suggests either $\beta_0 + \beta_2 < 0$ (low growth), or $\beta_0 + \beta_1 + \beta_2 + \beta_3 < 0$ (high growth), or both; and (ii) given X, V will be an increasing function of B in the low-profitability region, which suggests either $\beta_0 > 0$ (low growth), or $\beta_0 + \beta_1 > 0$ (high growth), or both. The estimates are (i) $\beta_0 + \beta_2 = -0.074$ (t = −2.39), and $\beta_0 + \beta_1 + \beta_2 + \beta_3 = -0.257$ (t = −2.53); and (ii) $\beta_0 = 0.569$ (t = 4.68) and $\beta_0 + \beta_1 = 0.687$ (t = 4.70). These coefficients all have the predicted signs and are significant at the 0.05 level or better.

H3 predicts that growth (i) will reduce the slope of the *V-B* relation, given X, in the high-profitability region, $\beta_1 + \beta_3 < 0$, and (ii) will increase the slope in the low-profitability region, $\beta_1 > 0$. The results show that $\beta_1 + \beta_3 = -0.182$ (t = −2.32) and $\beta_1 = 0.117$ (t = 4.15). Again, the coefficients have the predicted signs and are significant at the 0.1 level or better.

Together, the results of Regression (8) are consistent with all aspects of the growth effect predicted in H1, H2, and H3 holding jointly (as well as separately).

In addition, our model (8) displays the valuation properties documented in prior studies. For example, the result that the slope coefficient on X is greater in the high- than in the low-profitability region, both for low- and high-growth firms, implies the convexity of the *V-X* relation, holding B constant (recognizing that $q=X/B$). Similarly, the result that the slope on B is greater in the low- than in the high-profitability region, for both low- and high-growth firms, implies the convexity of the *V-B* relation, holding X constant. The convexity of these relations has been shown by Burgstahler and Dichev (1997). Furthermore, these same results also mean that the valuation emphasis shifts from earnings to equity book value as we move from high-profitability to low-profitability firms, consistent with the finding of Collins et al. (1999) from their analysis of profit and loss firms.

Testing H4

H4 requires that q stay constant as B varies within a sample. This means that X ($X = Bq$) must vary with B proportionally when q stays fixed. Because we cannot test H4 from (8) for a general q-value,[18] we construct a test of H4 at two representative values of q, namely, the mean q-values

[17] Again, tests here are based on Bonferroni-adjusted critical t-values. While H1, H2, and H3 together involve eight separate predictions (assuming that we test a strict version of H2, i.e., to test it separately for high and low growth), there are *six* independent restrictions, with the remaining two implied by the others. The tests are one-sided.

[18] The coefficient on B, given q, will be a function of the coefficients in Equation (8) and q. So without setting q to a specific value, neither this coefficient nor its t-value can be computed.

TABLE 8

Growth Effect in the Full Model (H1, H2, H3)

Variable	Coefficient	Predicted Sign	Estimate	t-value	VIF
Base Intercept			16.414[a]	10.50	
H_g			3.044[a]	14.99	2.158
H_q			3.046[a]	12.31	2.158
H_gH_q			1.034[a]	4.69	2.014
B [H2: slope in low g and low q]	β_0	+	0.569[a],***	4.68	4.761
H_gB [H3: growth effect in low q]	β_1	+	0.117[a],***	4.15	4.595
H_qB	β_2	−	−0.644[a]	−6.14	4.651
H_gH_qB	β_3	−	−0.299[a]	−5.31	4.751
X	γ_0	+	4.894[a]	9.89	5.093
H_gX [H1: growth effect in low q]	γ_1	+/−	−0.050	−0.19	5.067
H_qX	γ_2	+	6.776[a]	7.53	3.956
H_gH_qX [H1: incremental growth effect in high versus low q]	γ_3	+	1.860[a],***	3.46	5.084
$H_gX + H_gH_qX$ [H1: growth effect in high q]	$\gamma_1 + \gamma_3$	+	1.811[b],*	2.31	
$B + H_gB$ [H2: slope in high g and low q]	$\beta_0 + \beta_1$	+	0.687[a],***	4.70	
$B + H_qB$ [H2: slope in low g and high q]	$\beta_0 + \beta_2$	−	−0.074[a],**	−2.39	
$B + H_gB + H_qB + H_gH_qB$ [H2: slope in high g and high q]	$\beta_0 + \beta_1 + \beta_2 + \beta_3$	−	−0.257[a],**	−2.53	
$H_gB + H_gH_qB$ [H3: growth effect in high q]	$\beta_1 + \beta_3$	−	−0.182[b],*	−2.32	
Adj. R^2			0.63		

[a], [b] Indicate one-tailed significance at the 0.01 and 0.05 levels without Bonferroni corrections, respectively.
*, **, *** Indicate one-tailed significance at the 0.1, 0.05, and 0.01 levels with Bonferroni corrections (m = 6), respectively.

(continued on next page)

TABLE 8 (continued)

This table reports the results of Regression (8):

$$V_i = \alpha_0 + \alpha_1 H_g + \alpha_2 H_q + \alpha_3 H_g H_q + \beta_0 B_i + \beta_1 H_g B_i + \beta_2 H_q B_i + \beta_3 H_g H_q B_i + \gamma_0 X_i + \gamma_1 H_g X_i + \gamma_2 H_q X_i + \gamma_3 H_g H_q X_i + e_i,$$

where H_q is an indicator variable equal to 1 for firms in the high half of profitability in a sample, and H_g is an indicator variable equal to 1 for firms in the high half of growth. The explanatory factors that are continuous variables are mean-adjusted. We run regressions year by year, from 1966 to 2003. The reported results are the average coefficients across the years, and t-values are based on Fama-MacBeth standard errors with Newey-West adjustments.

in the low- and high-q regions. Table 9 presents the results.

At the mean q-value in the low-profitability region ($q = -0.090$), the coefficient on B, as implied by Regression (8), equals 0.980 (t = 14.88); this is the slope of the *V-B* relation for low-growth firms. At the same q-value, the slope increases to 1.298 (t = 12.47) for high-growth firms; the coefficient difference of 0.318 (t = 7.15) is significant at the 0.01 level.

The results are qualitatively the same in the high-profitability region. At the mean q-value in this region ($q = 0.229$), the coefficient on B is 2.089 (t = 21.42) for low-growth firms, and increases to 2.217 (t = 22.82) for high-growth firms; the difference of 0.128 (t = 3.07) is significant at the 0.01 level. These results are consistent with H4, which predicts that the slope of the *V-B* relation will be steeper for high-growth than for low-growth firms.

Economic Importance of the Growth Effect

To assess the economic importance of the growth effect, we address two specific issues in this subsection: (i) how much of a difference the growth variable makes in explaining variations in equity value induced by earnings or book value, and (ii) how important it is to condition the growth effect on profitability. As in our hypothesis testing, we focus on the growth effect conveyed through the slope coefficients on earnings and book value.

Economic Effect of Growth on Predicted Value Changes Induced by Earnings

We first illustrate the importance of the growth variable in predicting equity value changes induced by earnings, at a given book value (set at the overall sample mean, \$10.918). In the high-q region, the X-value corresponding to the 3rd quartile of q in the sample ($q = 0.190$) is \$2.074. We pick two earnings points at \$0.5 below and above this amount, that is, X = \$1.574 and \$2.574, respectively. Following the results of Regression (8), as X varies from \$1.574 to \$2.574 holding B

TABLE 9

Growth and the Relation between Equity Value and Book Value at Low and High Profitability Points (H4)

q-value		Predicted Sign	*B* Coefficient	t-value
$q_L = -0.090$	Slope in low g region	+	0.980[a],***	14.88
	Incremental slope in high g	+	0.318[a],***	7.15
	Slope in high g	+	1.298[a],***	12.47
$q_H = 0.229$	Slope in low g region	+	2.089[a],***	21.42
	Incremental slope in high g	+	0.128[a],***	3.07
	Slope in high g	+	2.217[a],***	22.82

[a] Indicates one-tailed significance at the 0.01 level without Bonferroni corrections.

*** Indicates one-tailed significance at the 0.01 level with Bonferroni corrections (m = 4).

This table reports results for H4 (at specific q-values) that are derived from Regression (8) (Table 8):

$$V_i = \alpha_0 + \alpha_1 H_g + \alpha_2 H_q + \alpha_3 H_g H_q + \beta_0 B_i + \beta_1 H_g B_i + \beta_2 H_q B_i + \beta_3 H_g H_q B_i + \gamma_0 X_i + \gamma_1 H_g X_i + \gamma_2 H_q X_i + \gamma_3 H_g H_q X_i + e_i.$$

We select the mean q-values in the low- and high-q groups, denoted q_L and q_H, respectively, and determine the slope of the *V-B* relation at these points. For the low-profitability group, the slope of B is $\beta_0 + \gamma_0 q_L$ for low-growth firms, and $\beta_0 + \beta_1 + (\gamma_0 + \gamma_1) q_L$ for high-growth firms. For the high-profitability group, the slope of B is $\beta_0 + \beta_2 + (\gamma_0 + \gamma_2)\, q_H$ for low-growth firms, and $\beta_0 + \beta_1 + \beta_2 + \beta_3 + (\gamma_0 + \gamma_1 + \gamma_2 + \gamma_3) q_H$ for high-growth firms. The results are averaged across the years, and t-values are based on Fama-MacBeth standard errors with Newey-West adjustments.

at \$10.918, the predicted V-value moves from \$25.517 to \$37.187 for low-growth firms ($H_g = 0$), an increase of \$11.67 (=37.187 − 25.517).[19] For the same variation in X, the predicted V moves from \$30.534 to \$44.014 for high-growth firms ($H_g = 1$), an increase of \$13.48. The difference in the predicted V-change between low- and high-growth firms is \$1.81, which is about 5 percent of the value of the underlying firms, an economically significant amount. For reference, the profitability levels associated with the two earnings points above are 0.144 and 0.236, respectively, a difference of 0.092, which is relatively small as compared to the standard deviation of 0.289 in the overall sample.

In the low-q region, we choose X-values at \$0.5 below and above the 1st q-quartile (q = 0.019), that is, X = −\$0.293 and \$0.707, respectively. Following the results of Regression (8), as X varies from–\$0.293 to \$0.707 holding B at \$10.918, the predicted V moves from \$9.817 to \$14.711 in the low-growth group, an increase of \$4.894. For the same variation in earnings, the predicted equity value moves from \$12.928 to \$17.772 in the high-growth group, an increase of \$4.844, about the same as in the low-growth group. The difference between these two predicted value changes of–\$0.05 is less than 0.4 percent of the value of the underlying firms. Thus, in low-profitability regions, growth makes almost no difference economically in predicting equity value changes induced by earnings. The contrasting results between high- and low-profitability firms highlight the role of profitability in determining the growth effect in valuation, as H1 predicts.

Economic Effect of Growth on Predicted Value Changes Induced by Equity Book Value

We similarly illustrate the growth effect on predicted value changes induced by book value, given earnings. In the high-q region, we set X at \$2.0 and choose B at \$3 below and \$3 above the sample mean of \$10.918; that is, B= \$7.918 and \$13.918, respectively. The corresponding q-values are 0.253 and 0.144. Among low-growth firms, as B moves from \$7.918 to \$13.918 holding X at \$2.0, the predicted V decreases from \$30.713 to \$30.263, a small reduction of \$0.45. Among high-growth firms, the same change in B causes the predicted V to decrease from \$37.048 to \$35.505, a more significant reduction of \$1.543. The *difference* in the predicted V-change between the high- and low-growth firms is \$1.093, about 3 percent of the underlying equity value, which is economically important.

In the low-q region, we set X at –\$0.5 (loss firms) and again choose B at \$3 below and above the sample mean; that is, B = \$7.918 and \$13.918, respectively. The corresponding q-values are −0.063 and −0.036. Among low-growth firms, as B moves from \$7.918 to \$13.918 holding X at –\$0.5, the predicted V changes from \$7.097 to \$10.511, an increase of \$3.414. On the other hand, among high-growth firms, the same variation in B causes the predicted V to change from \$9.868 to \$13.984, an increase of \$4.116. In this case, the *difference* in the predicted V-change between high- and low-growth firms is \$0.702, about 7 percent of the underlying firms' equity value, an economically significant amount.

To summarize, in the high-q region increases in equity book value cause faster equity value declines among high- (versus low-) growth firms, whereas in the low-q region, increases in equity book value cause equity value to rise faster among high- (versus low-) growth firms.

---

[19] In this subsection, all equity values are calculated based on the estimated Equation (8) (reported in Table 8): $V = 16.414 + 3.044H_g + 3.046H_q + 1.034H_gH_q + 0.569B' + 0.117H_gB' - 0.644H_qB' - 0.299H_gH_qB' + 4.894X' - 0.05H_gX' + 6.776H_qX' + 1.86H_gH_qX'$, where $X' = X - 1.055$ and $B' = B - 10.918$ are mean-adjusted earnings and equity book value, respectively. For example, the predicted V at X = 1.574 conditional on B = 10.918, $H_g = 0$, and $H_q = 1$ is V = 16.414+3.046+ (0.569 − 0.644) × (10.918 − 10.918) + (4.894 + 6.776) × (1.574 − 1.055) = 25.517; and that at X = 2.574 is V = 16.414 + 3.046 + (0.569 − 0.644) × (10.918 − 10.918) + (4.894 + 6.776) × (2.574 − 1.055) = 37.187.

VI. CONSERVATISM, PAST INVESTMENT ACTIVITIES, AND THE RELATION BETWEEN EQUITY VALUE AND ACCOUNTING DATA

To test H5 concerning the role of past investments that arises from accounting conservatism, we split annual samples into halves based on the rate of past investment increases, measured by the average annual change in capital expenditures scaled by the year-beginning equity book value over the past three years (the further data requirement reduces the sample to 76,662 observations).[20] Following H5, firms with faster (slower) investment increases have high (low) earnings conservatism, labeled *Hc* (*Lc*). The following regression, generalized from (8), estimates a separate value-accounting relation for the *Lc* and *Hc* groups:

$$\begin{aligned} V_i = L_c\{&\alpha_{0L} + \alpha_{1L}H_g + \alpha_{2L}H_q + \alpha_{3L}H_gH_q + \beta_{0L}B_i + \beta_{1L}H_gB_i + \beta_{2L}H_qB_i + \beta_{3L}H_gH_qB_i + \gamma_{0L}X_i \\ &+ \gamma_{1L}H_gX_i + \gamma_{2L}H_qX_i + \gamma_{3L}H_gH_qX_i\} + H_c\{\alpha_{0H} + \alpha_{1H}H_g + \alpha_{2H}H_q + \alpha_{3H}H_gH_q + \beta_{0H}B_i \\ &+ \beta_{1H}H_gB_i + \beta_{2H}H_qB_i + \beta_{3H}H_gH_qB_i + \gamma_{0H}X_i + \gamma_{1H}H_gX_i + \gamma_{2H}H_qX_i + \gamma_{3H}H_gH_qX_i\} + e_i. \quad (9) \end{aligned}$$

In (9), the three-way sorting by conservatism, growth, and profitability is performed independently of one another. The purpose of (9) is twofold: (i) to examine the validity of the predicted growth effect (H1, H2, and H3) in a more refined setting that differentiates the high-earnings-conservatism group from the low group, and (ii) to examine the role of earnings conservatism (H5) in a regression that incorporates the growth effect. Table 10 reports the results of Regression (9), as explained below.

Growth Effect in the Low-Conservatism Group

Table 10, Panel A reports the results on the growth effect in the low-conservatism group. In the high-profitability region, the effect of high (versus low) investment growth on the slope of the *V-X* relation given *B*, $\gamma_{1L} + \gamma_{3L}$, has a value of 2.917 (t = 3.52), significant at the 0.01 level.[21] This effect decreases to 0.386 (t = 0.82) in the low-profitability region. The difference between them (γ_{3L}) of 2.531 (t = 6.82) is significant at the 0.01 level, consistent with H1.

In the high-profitability region, the slope of the *V-B* relation given *X* is represented by $\beta_{0L} + \beta_{2L}$ for low-growth firms, which has a negative value of -0.187 (t = -5.50), whereas the slope is represented by $\beta_{0L} + \beta_{1L} + \beta_{2L} + \beta_{3L}$ for high-growth firms, which has an even more negative value of -0.506 (t = -7.21). The difference between them ($\beta_{1L} + \beta_{3L}$) of -0.319 (t = -3.39) is significant at the 0.01 level.

In contrast, in the low-profitability region, the slope of the *V-B* relation for low-growth firms (β_{0L}) has a positive value of 0.528 (t = 4.05), and that for high-growth firms ($\beta_{0L} + \beta_{1L}$) is 0.538 (t = 3.17), consistent with H2. The slope difference between high- and low-growth firms (β_{1L}) of 0.010 (t = 0.25) is statistically insignificant. These results are consistent with H2 and H3(i), but inconsistent with H3(ii).

Overall, the above results are consistent with all parts of the growth effect, stated in H1, H2, and H3, that are predicted by Zhang's (2000) model. However, we find no evidence for the intuitive prediction in part (ii) of H3.

---

[20] In previous sections where we examine the effect of investment growth, what is a proper measure of investment is less clear-cut. There, we use changes in equity book value to proxy for investment (with alternative proxies used in robustness checks). This can be justified on the ground that financial (as well as operating) assets are part of investment because they generate earnings themselves and can be used to fund future operations. But in this section, "investment" is more clearly defined in Zhang (2000), which refers to assets depreciated over multiple periods. It follows that, here, capital expenditures better proxy for investment than alternatives such as equity book value changes.

[21] Again, we apply Bonferroni corrections to critical t-values here, recognizing all the conditions imposed by H1, H2, and H3, and requiring them to hold in both the low- and high-conservatism groups.

TABLE 10

Conservatism and Growth Effects in the Full Model (H1, H2, H3, H5)

Panel A: Low Earnings Conservatism (L_c) Group

Variable	Coefficient	Predicted Sign	Estimate	t-value	VIF
Base Intercept			15.113[a]	8.66	
H_g			1.679[a]	9.02	2.503
H_q			−0.312	−0.82	2.910
H_gH_q			−0.990[a]	−3.21	2.721
B [H2: slope in low g and low q]	β_{0L}	+	0.528[a],***	4.05	4.562
H_gB [H3: growth effect in low q]	β_{1L}	+	0.010	0.25	4.300
H_qB	β_{2L}	−	−0.715[a]	−4.53	5.120
H_gH_qB	β_{3L}	−	−0.329[a]	−5.92	5.127
X	γ_{0L}	+	5.993[a]	13.28	5.520
H_gX [H1: growth effect in low q]	γ_{1L}	+/−	0.386	0.82	5.496
H_qX	γ_{2L}	+	7.542[a]	4.84	5.030
H_gH_qX [H1: incremental growth effect in high versus low q]	γ_{3L}	+	2.531[a],***	6.82	6.620
$H_gX + H_gH_qX$ [H1: growth effect in high q]	$\gamma_{1L} + \gamma_{3L}$	+	2.917[a],***	3.52	
$B + H_gB$ [H2: slope in high g and low q]	$\beta_{0L} + \beta_{1L}$	+	0.538[a],***	3.17	
$B + H_qB$ [H2: slope in low g and high q]	$\beta_{0L} + \beta_{2L}$	−	−0.187[a],***	−5.50	
$B + H_gB + H_qB + H_gH_qB$ [H2: slope in high g and high q]	$\beta_{0L} + \beta_{1L} + \beta_{2L} + \beta_{3L}$	−	−0.506[a],***	−7.21	
$H_gB + H_gH_qB$ [H3: growth effect in high q]	$\beta_{1L} + \beta_{3L}$	−	−0.319[a],***	−3.39	

Panel B: High Earnings Conservatism (H_c) Group

Variable	Coefficient	Predicted Sign	Estimate	t-value	VIF
Base Intercept			19.159[a]	10.06	
H_g			2.861[a]	17.05	2.218

(continued on next page)

Panel B: High Earnings Conservatism (H_c) Group

Variable	Coefficient	Predicted Sign	Estimate	t-value	VIF
H_q			2.943[a]	12.64	2.422
H_gH_q			0.906[a]	3.02	2.223
B [H2: slope in low g and low q]	β_{0H}	+	0.426[a],**	3.04	4.764
H_gB [H3: growth effect in low q]	β_{1H}	+	−0.044	−0.87	4.799
H_qB	β_{2H}	−	−0.799[a]	−5.36	4.561
H_gH_qB	β_{3H}	−	−0.415[a]	−9.27	4.749
X	γ_{0H}	+	6.895[a]	11.76	5.902
H_gX [H1: growth effect in low q]	γ_{1H}	+/−	0.466	0.99	5.942
H_qX	γ_{2H}	+	7.827[a]	5.31	4.538
H_gH_qX [H1: incremental growth effect in high versus low q]	γ_{3H}	+	2.552[a],***	8.00	5.836
$H_gX + H_gH_qX$ [H1: growth effect in high q]	$\gamma_{1H} + \gamma_{3H}$	+	3.018[a],***	5.45	
$B + H_gB$ [H2: slope in high g and low q]	$\beta_{0H} + \beta_{1H}$	+	0.382[a],***	4.71	
$B + H_qB$ [H2: slope in low g and high q]	$\beta_{0H} + \beta_{2H}$	−	−0.373[a],***	−13.10	
$B + H_gB + H_qB + H_gH_qB$ [H2: slope in high g and high q]	$\beta_{0H} + \beta_{1H} + \beta_{2H} + \beta_{3H}$	−	−0.833[a],***	−10.44	
$H_gB + H_gH_qB$ [H3: growth effect in high q]	$\beta_{1H} + \beta_{3H}$	−	−0.459[a],***	−3.93	
Adj. R^2			0.87		

Panel C: Testing H5 (comparing earnings coefficients between H_c and L_c)

Variable	Predicted Sign	Estimate	t-value
Diff. in X coeff. in (low g, low q): $H_c - L_c$	+	0.902[a],***	3.46
Diff. in X coeff. in (high g, low q): $H_c - L_c$	+	0.982[a],***	3.16
Diff. in X coeff. in (low g, high q): $H_c - L_c$	+	1.188[a],**	2.56
Diff. in X coeff. in (high g, high q): $H_c - L_c$	+	1.289[a],***	3.31

[a] Indicates one-tailed significance at the 0.01 level without Bonferroni corrections.
, * Indicate one-tailed significance at the 0.05 and 0.01 levels with Bonferroni corrections (m = 12 for testing the growth effect and m = 4 for testing the conservatism effect), respectively.

(continued on next page)

This table reports the results of Regression (9):

$$V_i = L_c\{\alpha_{0L} + \alpha_{1L}H_g + \alpha_{2L}H_q + \alpha_{3L}H_gH_q + \beta_{0L}B_i + \beta_{1L}H_gB_i + \beta_{2L}H_qB_i + \beta_{3L}H_gH_qB_i + \gamma_{0L}X_i + \gamma_{1L}H_gX_i + \gamma_{2L}H_qX_i + \gamma_{3L}H_gH_qX_i\} + H_c\{\alpha_{0H} + \alpha_{1H}H_g + \alpha_{2H}H_q + \alpha_{3H}H_gH_q + \beta_{0H}B_i$$
$$+ \beta_{1H}H_gB_i + \beta_{2H}H_qB_i + \beta_{3H}H_gH_qB_i + \gamma_{0H}X_i + \gamma_{1H}H_gX_i + \gamma_{2H}H_qX_i + \gamma_{3H}H_gH_qX_i\} + e_i,$$

where $L_c(H_c)$ indicates firms with low (high) earnings conservatism, which are firms having average investment increases over the past three years below (above) the sample median, with investment increases measured by changes in capital expenditures scaled by equity book value. H_q is an indicator variable equal to 1 for firms in the high half of profitability in a sample, with profitability defined as earnings per share divided by equity book value per share at the beginning of the year. H_g is an indicator variable equal to 1 for firms in the high half of growth, with growth measured as the average annual growth rate of equity book value in the subsequent three years. We run regressions year by year, from 1968 to 2003. The reported results are the average coefficients across the years, and t-values are based on Fama-MacBeth standard errors with Newey-West adjustments.

Growth Effect in the High-Conservatism Group

The same qualitative properties hold in the high-earnings-conservatism group, as shown in Table 10, Panel B. In the high-profitability region, the effect of high (versus low) growth on the slope of the *V-X* relation ($\gamma_{1H}+\gamma_{3H}$) is 3.018 (t = 5.45); this effect decreases to 0.466 (t = 0.99) in the low-profitability region. The difference (γ_{3H}) is 2.552 (t = 8.00), significant at the 0.01 level.

In the high-profitability region, the *V-B* relation given *X* has a negative slope ($\beta_{0H}+\beta_{2H}$) equal to −0.373 (t = −13.10) for low-growth firms, which turns even more negative for high-growth firms ($\beta_{0H}+\beta_{1H}+\beta_{2H}+\beta_{3H}$), equal to −0.833 (t = −10.44). The difference in slope between them ($\beta_{1H}+\beta_{3H}$) is −0.459 (t = − 3.93), significant at the 0.01 level. On the other hand, in the low-profitability region, the *V-B* relation given *X* has a positive slope both for low-growth firms (β_{0H}), equal to 0.426 (t = 3.04), and for high-growth firms ($\beta_{0H}+\beta_{1H}$), equal to 0.382 (t = 4.71). The slope difference between them (β_{1H}) of −0.044 (t = −0.87) is insignificant.

Thus, as in the low-conservatism group, the results are consistent with all parts of the growth effect stated in H1, H2, and H3 that are based on theoretical predictions, but no evidence is found for the intuitive prediction in part (ii) of H3.

Conservatism and the Valuation Impact of Earnings (H5)

H5 predicts that firms that have just experienced rapid increases in investments will have their earnings more conservatively stated than do those that experienced slow (or negative) increases in recent investments. Thus, *ceteris paribus*, the earnings coefficient should be greater in the former than in the latter group. Regression (9) contains four different combinations of *g* and *q* (as measured by their respective indicator variables); a separate earnings coefficient is estimated for each combination. Table 10, Panel C shows that the earnings coefficient is greater in the high- than in the low-conservatism group in all four scenarios identified by (*g*, *q*) combinations. The coefficient difference between the high- and low-conservatism groups is 0.902 (t = 3.46) in scenario (*Low g, Low q*); 0.982 (t = 3.16) in (*High g, Low q*); 1.188 (t = 2.56) in (*Low g, High q*); and 1.289 (t = 3.31) in (*High g, High q*). These differences are significant at the 0.05 level or better, consistent with H5.[22]

VII. ROBUSTNESS CHECKS

We also perform a range of robustness checks employing alternative variable measures and regression specifications, and find the qualitative results generally unchanged.

In our main analysis, growth is measured by average *ex post* growth in equity book value. Alternative measures we use are *ex post* growth in total assets, *ex post* growth in operating assets, and *ex post* capital expenditures (scaled by property, plant, and equipment). In addition to using a simple high-low indicator variable to measure growth, we also use the percentile rankings of growth in our tests and obtain similar results.

Beyond these *ex post* growth proxies, we also use the median analyst forecast of long-term earnings growth in an industry to proxy for the expected investment growth of firms in that industry. This *ex ante* measure mitigates the concern of survivorship biases potentially introduced by *ex post* proxies. While we have adjusted growth for inflation (the Consumer Price Index) in the main analysis, we also conduct an analysis without this adjustment.

In our tests using partial model specifications (Section IV), we control for a variable by the method of scaling. Alternatively, we also use the method of partitioning, whereby we partition

[22] For each given (*g*, *q*)-combination, we find generally little difference in mean (median) *q*- and g-values between the L_c and H_c groups.

annual samples along the dimension of the control variable (to reduce the extent of its variation) and then run regressions separately for the partitions. With this alternative method, we are able to include negative- as well as positive-earnings observations in the tests. Our results are also consistent with the predicted growth effect.

The theoretical model underlying our study assumes the clean surplus relation. As Ohlson (2005) points out, in situations in which a firm engages in equity transactions with investors that change the number of shares outstanding, the clean surplus relation does not hold on a per-share basis, and should be applied to the whole firm (using total equity value, book value, and earnings). To mitigate this concern, we run two additional sets of regressions. In one set, we use total equity value, equity book value, and earnings instead of per-share amounts. In the second set, we restrict the sample to those firms that experience a change of less than 1 percent in the number of shares outstanding between the preceding and the subsequent year (and so the clean surplus relation approximately holds on a per-share basis). In both sets of analysis, the results are similar to those reported above.

In Section VI, we proxy past investment by capital expenditures scaled by equity book value. We alternatively use PPE and total assets as scalars and find similar results.

VIII. SUMMARY AND CONCLUSIONS

This study incorporates the role of investment growth to reexamine the relation between equity value, earnings, and equity book value. Following the valuation model of Zhang (2000), we predict how investment growth, in conjunction with profitability, influences this relation. We then empirically test the predictions and find consistent results. In addition to exploring the valuation effect of prospective investments, which arises from economic incentives for value creation, this study also demonstrates how past investments in conjunction with conservative accounting practices affect the valuation role of earnings. Our results provide insight into the important role of investment activities in mediating the value-accounting relation.

Valuation research is of interest to investors, standard-setters, and academic researchers. In practice, investors and analysts adopt different valuation techniques and emphasize various fundamental factors in financial forecasting. By presenting a more complete view of the relation between equity value and accounting data, our study provides a framework for thinking about how to perform financial analysis and valuation. In particular, it explains that both prospective investment growth and past investment activities alter the way accounting data are mapped into value. Relative valuation techniques, such as price-to-earnings and price-to-book multiples, are widely used in investment and business decisions. By demonstrating how the relation between equity value and earnings, and similarly that between equity value and book value, depends on investment growth, the study gives a better understanding of the fundamental determinants of these multiples. This is useful not only for identifying benchmark firms in multiples-based valuations, but also for determining the appropriate levels for the multiples. Likewise, for standard-setters who are responsible for designing reporting rules to enhance the usefulness of financial data, a more in-depth knowledge of the link between accounting data and value will be beneficial for formulating more targeted standards.

In accounting research, parsimonious valuation models are often used in which earnings and equity book value serve as explanatory variables for equity value. However, the circumstances in which such simplified models are appropriate are not always well understood. Our study shows that the relation between equity value and accounting data is more complex than what has been previously documented or assumed in the literature; in particular, growth is an important attribute that intervenes in this relation. The study also shows that in cross-sections, the behavior of equity

value in relation to a specific accounting variable (such as equity book value) can differ substantially along different dimensions, such that it is plausible to obtain different, or even opposite, results from different empirical samples.

Finally, our results have implications for research that applies valuation models to address accounting-related issues. Given our finding that the valuation impacts of earnings and book value depend on investment activities, one needs to be cautious when drawing inferences concerning, for example, the value relevance of accounting data or the degree of accounting conservatism. In such studies, researchers should separate the effect of a firm's economic condition from that of accounting.

APPENDIX
MATHEMATICAL DERIVATIONS OF THE GROWTH EFFECT

This appendix presents the mathematical derivations to show how investment growth affects the relation between equity value and accounting variables (H1, H3, and H4). Note that H2 was derived in Zhang (2000, 284–285).

Growth and the Relation between Equity Value and Earnings, Given Book Value (H1)

Differentiating V given by model (1) with respect to X, holding B constant, we get $\frac{dV}{dX}|_B = P'(q) + k + gC'(q) > 0$. Differentiating $\frac{dV}{dX}|_B$ with respect to g yields $d(\frac{dV}{dX}|_B) / dg = C'(q) > 0$. H1 follows by further noting that $C'(q)$ increases with q.

Growth and the Relation between Equity Value and Book Value, Given Earnings (H3)

Differentiating V with respect to B, given X, and simplifying, we have $\frac{dV}{dB}|_X = -qP'(q) + P(q) - gqC'(q) + gC(q)$. Differentiating $\frac{dV}{dB}|_X$ with respect to g yields $d(\frac{dV}{dB}|_X) / dg = -qC'(q) + C(q)$, which is negative when q is sufficiently large to make the growth option valuable. This leads to part (i) of H3, relevant to high-profitability regions. The model does not address the case of high growth by low-profitability firms, as related to part (ii) of H3.

Growth and the Relation between Equity Value and Book Value, Given Profitability (H4)

Differentiating V with respect to B, given q, we get $\frac{dV}{dB}|_q = P(q) + kq + gC(q) > 0$, indicating that the marginal valuation impact of book value, given profitability, is a function of q and g. Differentiating $\frac{dV}{dB}|_q$ with respect to g, we get $d(\frac{dV}{dB}|_q) / dg = C(q) > 0$. Part (i) of H4 follows. Again, the model does not address the case of high growth by low-profitability firms, as related to part (ii) of H4.

REFERENCES

Aiken, L. S., and S. G. West. 1991. *Multiple Regression: Testing and Interpreting Interactions*. Newbury Park, CA: Sage Publications.

Bar-Yosef, S., J. Callen, and J. Livnat. 1987. Autoregressive modeling of earnings investment causality. *The Journal of Finance* 42: 11–28.

Barth, M. 1991. Relative measurement errors among alternative pension asset and liability measures. *The Accounting Review* 66: 433–463.

———, W. Beaver, and W. Landsman. 1992. The market valuation implications of net periodic pension cost components. *Journal of Accounting and Economics* 15 (1): 27–62.

Biddle, G., P. Chen, and G. Zhang. 2001. When capital follows profitability: Nonlinear residual income dynamics. *Review of Accounting Studies* 6 (2–3): 229–265.

Burgstahler, D., and I. Dichev. 1997. Earnings, adaptation, and equity value. *The Accounting Review* 73: 187–215.

Callen, J., and D. Segal. 2005. Empirical tests of the Feltham-Ohlson (1995) model. *Review of Accounting Studies* 10 (4): 409–429.

Chen, P., and G. Zhang. 2007. How do accounting variables explain stock price movements? Theory and evidence. *Journal of Accounting and Economics* 43 (2–3): 219–244.

Collins, D., and S. P. Kothari. 1989. An analysis of the cross-sectional and intertemporal determinants of earnings response coefficients. *Journal of Accounting and Economics* 11 (2–3): 143–181.

———, M. Pincus, and H. Xie. 1999. Equity valuation and negative earnings: The role of book value of equity. *The Accounting Review* 74 (1): 29–61.

Core, J., W. Guay, and A. Van Buskirk. 2003. Market valuations in the new economy: An investigation of what has changed. *Journal of Accounting and Economics* 34 (1–3): 43–67.

Dechow, P., A. Hutton, and R. Sloan. 1999. An empirical assessment of the residual income valuation model. *Journal of Accounting and Economics* 26 (1–3): 1–34.

Dixit, A., and R. Pindyck. 1994. *Investment under Uncertainty*. Princeton, NJ: Princeton University Press.

Fama, E., and M. Miller. 1972. *The Theory of Finance*. Hinsdale, IL: Dryden Press.

Feltham, G., and J. Ohlson. 1995. Valuation and clean surplus accounting for operating and financial activities. *Contemporary Accounting Research* 11 (2): 689–731.

Holthausen, R., and R. Watts. 2001. The relevance of the value-relevance literature for financial accounting standard setting. *Journal of Accounting and Economics* 31 (1–3): 3–75.

Kothari, S. P. 2001. Capital markets research in accounting. *Journal of Accounting and Economics* 31 (1–3): 105–231.

Landsman, W. 1986. An empirical investigation of pension fund property rights. *The Accounting Review* 61: 662–692.

Lev, B., and P. Zarowin. 1999. The boundaries of financial reporting and how to extend them. *Journal of Accounting Research* 37: 353–385.

Lo, K., and T. Lys. 2000. The Ohlson model: Contribution to valuation theory, limitations, and empirical applications. *Journal of Accounting, Auditing & Finance* 14: 147–161.

Miller, M., and F. Modigliani. 1961. Dividend policy, growth, and the valuation of shares. *The Journal of Business* 34 (4): 411–433.

Ohlson, J. 1995. Earnings, book value, and dividends in equity valuation. *Contemporary Accounting Research* 11 (2): 661–687.

———. 2005. On accounting-based valuation formulae. *Review of Accounting Studies* 10 (2–3): 323–347.

———, and B. Juettner-Nauroth. 2005. Expected EPS and EPS growth as determinants of value. *Review of Accounting Studies* 10 (2–3): 349–365.

Penman, S., and X. Zhang. 2002. Accounting conservatism, the quality of earnings, and stock returns. *The Accounting Review* 77 (2): 237–264.

Rajan, M., S. Reichelstein, and M. Soliman. 2007. Conservatism, growth, and return on investment. *Review of Accounting Studies* 1 (2–3): 325–370.

Shevlin, T. 1991. The valuation of R&D firms with R&D limited partnership. *The Accounting Review* 66: 1–21.

Sirri, E., and P. Tufano. 1998. Costly search and mutual fund flows. *The Journal of Finance* 53 (5): 1589–1622.

Zhang, G. 2000. Accounting information, capital investment decisions, and equity valuation: Theory and empirical implications. *Journal of Accounting Research* 38: 271–295.

THE ACCOUNTING REVIEW
Vol. 86, No. 2
2011
pp. 637–667

American Accounting Association
DOI: 10.2308/accr.00000022

Voluntary Disclosure to Influence Investor Reactions to Merger Announcements: An Examination of Conference Calls

Michael D. Kimbrough
University of Maryland

Henock Louis
The Pennsylvania State University

ABSTRACT: We find that bidders are more likely to hold conference calls at merger announcements when the mergers are financed with stock and when the transactions are large. After controlling for endogeneity, we also find that conference calls are associated with more favorable market reactions to merger announcements. A content analysis of merger-related information releases for a limited subsample indicates that the more favorable reaction is related to the fact that, compared to press releases, conference calls provide a greater volume of information and place greater emphasis on forward-looking details. We find no evidence that the superior announcement returns associated with conference calls subsequently reverse or that conference calls are positively associated with pre-merger announcement abnormal accruals. Overall, the results suggest that managers use conference calls around merger announcements to credibly convey favorable private information to the market.

Keywords: *conference calls; voluntary disclosure; mergers.*

JEL Classifications: *D82; G34; M41.*

I. INTRODUCTION

We examine the determinants and consequences of acquirers' decisions to supplement merger announcement press releases with conference calls. Our analysis of disclosure choices concurrent with merger announcements is based on the potentially important role of disclosure in influencing investors' responses to merger announcements. For example, Wayne Moore, former president of Goldman Sachs, argues:

We appreciate the helpful comments of two anonymous reviewers, Shelley Curling, Paul Fischer, Rebecca Hann, John Hassell, Frank Heflin, Steve Huddart, Steven Kachelmeier (senior editor), Bob Kaplan, Craig Lefanowicz, Thomas Lys, Greg Miller, Alexander Nekrasov, Mark Penno, Eddie Riedl, Jack Stecher, Siew Hong Teoh, Beverly Walther (editor), Ross Watts, Joanna Wu, Susan Watts, Jerry Zimmerman, and workshop participants at The George Washington University, Harvard Business School, University of Maryland, The Ohio State University, The Pennsylvania State University, Purdue University, Rice University, University of Rochester, University of Minnesota Empirical Accounting Conference, University of Virginia, and Washington University. We are grateful for outstanding research assistance provided by Leila Jamal, Susanna Kim, Luke Watson, James Zeitler, and Jane Zhang.

Editor's note: Accepted by Beverly Walther.

Submitted: June 2009
Accepted: July 2010
Published Online: March 2011

> It is critical that the announcement of a transaction be well received and that the bidder's stock reacts favorably. As a result, the time, effort, and care that goes into announcing a deal has increased significantly. And the content—the description of the strategic rationale and the quantification of the synergies and future earnings effects—has as well. (Moore et al. 1998, 12)

Lipin and Sirower (2003) maintain that effectively communicating a merger plan to investors involves disclosing specific and verifiable details about the integration plans, financial projections (including whether the merger is accretive or dilutive to earnings), the management team type, key assumptions underlying the anticipated success, and the basis for the purchase price. The interactive nature of conference calls provides managers with an ideal forum for communicating the kind of contextual information that Lipin and Sirower (2003) argue is necessary to gain investor confidence. A detailed content analysis of a limited subsample of merger announcement press releases and conference call transcripts confirms that bidders that hold conference calls at the time of their merger announcements provide a greater volume of disclosure and emphasize forward-looking details to a greater degree than bidders that rely solely on merger announcement press releases to introduce the proposed deals to the market.

While we expect managers in general to have incentives to effectively communicate the rationale for their proposed transactions due to the well-documented cost of capital benefits of forthcoming disclosure, we examine the circumstances in which these incentives are particularly pronounced in the merger and acquisition (M&A) context. Specifically, we expect disclosure incentives to be more pronounced for stock-based mergers because the costs of a stock-based merger and the likelihood of its completion are directly tied to the post-announcement value of the bidder's stock. Moreover, there is a general presumption that stock-for-stock mergers are often motivated by overvaluation as opposed to genuine business considerations (Shleifer and Vishny 2003). Therefore, managers who have valid economic rationales to undertake a stock-for-stock merger have incentives to provide additional information to investors to counteract the otherwise adverse market reaction. Consistent with these arguments, we find that conference calls are more likely for stock-for-stock mergers. We also expect acquirers' incentives to hold conference calls at merger announcements to be more pronounced for large and complex mergers where management's intent is more likely to be unclear to investors. Accordingly, we find that conference calls are more likely as the economic importance of the deal (as proxied by the deal value as a percentage of the acquirer's pre-announcement market value) increases.

We then examine the relation between management's decision to hold a conference call at the merger announcement and the initial market reaction to the announcement. After correcting for self-selection bias associated with the decision to hold merger-related conference calls using the inverse Mills' ratio, we find that bidders that hold conference calls at merger announcements experience announcement returns that are substantially higher than they would have experienced otherwise. We provide additional insights through a content analysis of merger-related information releases for a subsample of 20 transactions that are accompanied by conference calls and a matched sample of 20 economically similar transactions that are not accompanied by conference calls. This analysis indicates that the more favorable reaction is related to the fact that conference calls, compared to press releases, provide a greater volume of information and place greater emphasis on forward-looking details. We find no evidence that the superior announcement return subsequently reverses. Moreover, we find that stock-for-stock acquirers that hold conference calls concurrent with their merger announcements report less positive pre-announcement abnormal accruals than stock-for-stock acquirers that do not hold conference calls. This finding, which is consistent with Jo and Kim's (2007) evidence of a negative association between abnormal accruals and disclosure frequency prior to seasoned equity offerings (SEOs), indicates that acquirers' use of conference calls is not positively associated with opportunistic earnings management. Overall, the

evidence suggests that managers use conference calls around merger announcements to credibly convey genuine favorable private information to the market as opposed to hyping their stock prices.

Our study extends the empirical voluntary disclosure literature by demonstrating the usefulness of voluntary disclosure generally and conference calls specifically outside the context of routine accounting events. Traditionally, the empirical voluntary disclosure literature focuses on disclosure activity linked to predictable accounting releases, such as management forecasts made in advance of earnings announcements (e.g., Baginski and Hassell 1997) and expanded press releases and clarifying conference calls made at the time of earnings announcements (e.g., Francis et al. 2002; Bowen et al. 2002; Kimbrough 2005). By contrast, our research highlights the potential usefulness of voluntary disclosure in the context of a (relatively) infrequent but potentially important strategic decision. In this regard, our work is related to Leone et al. (2007), who link managers' disclosure decisions about the intended use of initial public offerings (IPOs) proceeds to the degree of IPO underpricing.

Our findings also add to the literature on how bidders respond to the incentives to maximize share prices around mergers. Other studies in this area have focused on pre-announcement strategies. Brockman and Martin (2009) suggest that firms engaging in stock-for-stock mergers issue optimistic earnings forecasts prior to the merger announcements. Erickson and Wang (1999) and Louis (2004) focus on pre-merger announcement earnings management. These pre-announcement strategies are likely to favorably influence an acquirer's share price *prior* to a merger announcement. However, the benefit bidders receive from pre-announcement earnings guidance or earnings management is limited because, as Erickson and Wang (1999) argue, upon the announcement of a stock-for-stock merger, rational investors are likely to understand the prior incentives of the bidding firm's managers and, therefore, are likely to attempt to reverse the portion of the pre-announcement stock valuation attributable to misleading earnings guidance and earnings management. In contrast to the extant studies, we focus on concurrent disclosure because a stock-for-stock acquirer must be particularly concerned with the market response to the merger proposal and the ensuing valuation effect, which ultimately determines the cost and the likelihood of successful completion of the merger.

Our study is also related to Lang and Lundholm (2000) and Jo and Kim (2007). Using a 1992 sample of 41 firms with an average market capitalization of less than \$84 million, Lang and Lundholm (2000) document an increase in disclosure six months prior to SEOs, but find no evidence that managers provide more forward-looking information prior to the SEOs. In addition, they find that firms that increase their pre-SEO disclosure activity also experience positive pre-SEO abnormal stock returns, but suffer much larger price declines at, and over the months after, the SEO announcement than a set of control firms. Jo and Kim (2007) also find that firms with market capitalization of less than \$150 million (between 1990 and 1997) and an "unsustained" increase in disclosure inflate earnings aggressively prior to SEOs and have more negative performance over the long run. These results are consistent with managers acting opportunistically by using voluntary disclosure to hype their stock prices prior to SEOs.[1]

Our study differs from Lang and Lundholm (2000) and Jo and Kim (2007) in that we examine managers' contemporaneous disclosures made specifically about the economic event for which the stock is being used. Also, while these studies focus on small firms, we focus on a broader sample with an average market capitalization of more than \$2 billion. Moreover, the interactive nature of

---

[1] We do not examine stock issuances because the Security Data Company (SDC) provides filing dates of stock issuances but not the announcement dates. An analysis of about 50 issuance announcements reveals that the two dates are often different. In addition, the existence of cash-financed mergers provides a natural control sample, a feature that does not exist in the SEO setting.

conference calls, as opposed to the static nature of press releases examined by Lang and Lundholm (2000) and Jo and Kim (2007), reduces managers' ability to mislead investors and limits potential misinterpretations by investors by forcing managers to respond to analyst inquiries.[2] Accordingly, in contrast to the extant studies, we find that: (1) acquirers provide substantial amounts of forward-looking information during merger-related conference calls, (2) acquirers that hold conference calls experience superior announcement returns, and (3) there is no evidence that the superior returns subsequently reverse. Thus, in contrast to Lang and Lundholm's (2000) results, our findings suggest that managers use contemporaneous disclosures made specifically about the economic event for which the stock is being used to convey genuine favorable private information.[3]

Section II discusses managers' motivation for holding conference calls at the merger announcement. Section III describes the sample. Section IV analyzes the determinants of merger-related conference calls. Section V analyzes the capital market effects of merger-related conference calls. Section VI analyzes whether acquirers use conference calls to hype their stock prices. Section VII analyzes the content of the press releases and conference call transcripts for selected transactions. Section VIII concludes.

II. THE MOTIVATION FOR HOLDING MERGER-RELATED CONFERENCE CALLS

The Incentives for Additional Disclosure at Merger Announcements

To the extent that a proposed merger represents a substantial change in business strategy, historical information is unlikely to satisfy investors' current demands for information. As Dye (1985) and Jung and Kwon (1988) argue, when the market is aware that managers have private information and managers choose not to share that information, the market infers that the undisclosed information is unfavorable. Because managers have private information about the rationales for proposed mergers and their intended benefits, their disclosure decisions could have a substantial impact on investors' reactions to merger announcements.

According to Lipin and Sirower (2003), an effective merger communication strategy: (1) provides a credible story with clear targets that can be communicated and accomplished by the bidder and monitored over time by investors, (2) removes uncertainty and provides direction to the organization, and (3) links post-merger integration plans to the economics of the transaction. They point to PepsiCo's successful communication strategy in relation to its $13.4 billion purchase of Quaker Oats in December 2001. They particularly note that "PepsiCo got off to a good start with a detailed press release and investor presentation, *supported by a lengthy analyst/investor call and a Web cast*" (Lipin and Sirower 2003, 24; emphasis added).

Although all bidders have incentives to provide supplementary disclosure due to the well-documented cost of capital and reputational benefits of forthcoming disclosure, the incentives are particularly strong for economically significant deals where the intensity of investor demand for supplemental information is pronounced. In addition, disclosure incentives are stronger for the average stock-for-stock bidder because of the direct impact of its stock price on a merger's cost

---

2 Our study does not depend upon an elaborate theory of which investor groups are targeted by conference calls. We presume that conference calls are used to favorably influence the collective response of investors to the merger announcement. Regulation Fair Disclosure was in effect for our entire sample; therefore, the information conveyed in conference calls is not limited to a subset of investors.

3 While we attribute the contrasting findings between our study and Lang and Lundholm's (2000) to differences between the two studies in the types of disclosures examined, it is important to note that gun-jumping laws that prevented disclosure of forward-looking information prior to the filing of a registration statement were in effect during the period studied by Lang and Lundholm (2000), whereas the Securities and Exchange Commission (SEC) has provided safe harbors for the ongoing release of forward-looking information during our sample period. However, this difference in regulatory regime is unlikely to fully account for the different results.

and likelihood of completion[4] and because, absent additional information, investors generally view the use of stock as a signal that managers have adverse private information (Travlos 1987).

Conference Calls versus Alternative Communication Means

Conference calls allow managers to increase the volume of information transmitted at the merger announcement and communicate in an interactive manner with numerous analysts at one time. Mayew and Venkatachalam (2009) also suggest that managers' affect, as reflected in their voice patterns, can provide useful cues about future performance and that investors at least partially respond to the cues provided by managerial voice patterns during conference calls. Therefore, conference calls provide an ideal forum for managers to communicate important merger details, such as the anticipated benefits of the merger and projections of merger-induced financial performance. Our focus on conference calls as our indicator of supplemental disclosure is based on past research that documents their informativeness to investors (e.g., Tasker 1998; Bowen et al. 2002; Kimbrough 2005). In Section VII, we perform a detailed content analysis of merger announcement press releases and conference call transcripts.

Firms typically announce their acquisitions through press releases. However, given the contextual nature of the information that must be disclosed as part of an effective communication strategy, it is unlikely that a merger press release will be sufficient for large and complex mergers, particularly when management intent is unclear. Acquirers also communicate through SEC filings; however, given the extant evidence that investors do not process SEC filings in a timely manner,[5] it is unlikely that managers would rely mainly on these filings to counteract the potential adverse signal conveyed by the announcement of a stock-for-stock merger. Acquirers also communicate through their websites. However, none of these alternative communication vehicles offers substitutes for the interactive feature of conference calls.

Acquirers may also engage in road show presentations subsequent to the announcement of a proposed transaction. If bidders that use conference calls also engage in road show disclosures, then it is possible that conference calls do not completely reflect the information (incremental to the press release) conveyed by management to investors. However, because we focus on managers' disclosure choices and investors' responses during the narrow merger announcement period, the existence of road shows is unlikely to affect our study's inferences insofar as road shows typically take place after the initial merger announcement period.[6] Furthermore, Regulation Fair Disclosure (Reg-FD) was in effect during our entire sample period.[7] Any bidder in our sample that

[4] Under fixed value stock-for-stock arrangements, the value of the consideration target shareholders receive is certain but the number of shares the target shareholders will receive depends on the bidder's pre-closing stock price. In this case, investor reaction to the merger announcement has a direct impact on the number of shares that the bidder must issue, thereby determining the amount of dilution of existing shareholders' voting power and share of profits. Under fixed exchange stock-for-stock arrangements, the number of bidding firm shares that the target shareholders will receive is certain but the value of those shares is uncertain. In this situation, the target shareholders face greater uncertainty about the value of the consideration they are receiving and are less likely to agree to the merger if the bidder's share price drops after the merger is announced. Also, many stock-based mergers have collar provisions allowing the proposed merger to be cancelled if the bidders' stock price falls below a certain range.

[5] In general, prior studies find no evidence that the market reacts to the information in the 8-K filings in a timely manner (Pastena 1979; Fried and Schiff 1981; Johnson and Lys 1990; Klock 1994; Schwartz and Soo 1995). Extant studies also find limited or no reaction to new information disclosed in 10-Q and 10-K filings (Foster et al. 1983; Cready and Mynatt 1991; Stice 1991; Easton and Zmijewski 1993; Chung et al. 2003; Griffin 2003; Hollie et al. 2005; Li and Ramesh 2009).

[6] As discussed in Section IV, we focus on the initial merger announcement period based on the expectation that the incentive to engage in disclosure to favorably affect the market's reaction is likely to be most pronounced at the time of the initial merger announcement.

[7] The SEC has strongly suggested that companies take the following steps when publicly disclosing material information (earnings, for example) in order to comply with Reg-FD. "First, issue a press release, distributed through regular channels, containing the information; Second, provide adequate notice, by a press release and/or website posting, of a

engaged in road show presentations would have had to disclose in advance to the public, through a press release or through a previously announced conference call, any material information that it planned to convey during the presentations. Our content analysis of press releases and conference calls in Section VII indicates that conference calls are the most likely forum for the release of information that management plans to subsequently disclose in road show presentations. Therefore, it seems reasonable to assume that conference calls capture the material information subsequently disclosed in road shows.

III. SAMPLE SELECTION AND DESCRIPTIVE STATISTICS

The study covers merger bids on Security Data Company's (SDC) online database of domestic mergers and acquisitions that were announced between January 1, 2002 and December 31, 2006. A transaction is included in the sample if it satisfies the following criteria: (1) the bidder is a publicly traded company; (2) the transaction value, the method of payment, and the merger announcement date are contained in SDC; (3) the bidder has necessary data on the CRSP/Compustat merged database to compute market capitalization, leverage, and book-to-market ratio; (4) the transaction value is at least 10 percent of the bidder's market value; (5) the merger announcement press release could be located in Factiva or on the bidder's website; and (6) the transaction was not completed at the time of the announcement.

Because hand-collection and review of merger announcement press releases is costly, we impose a minimum deal ratio requirement (i.e., requirement 4) and focus our efforts on economically important mergers, where the role for voluntary disclosure is likely to be particularly strong. We require access to the merger announcement press release (i.e., requirement 5) for several reasons. First, we refer to the press release to confirm key transactional data obtained from SDC on the method of payment and the total transaction value. Second, and more importantly, we rely on the press release to identify the timing of a conference call relative to the merger announcement. Our sample period is subsequent to the passage of Reg-FD, which prohibits selective disclosure of material information. U.S. firms have generally complied with this requirement by making their conference calls open to the public and by announcing upcoming conference calls in their press releases. Therefore, we examine the merger announcement press release for any announcement of merger-related conference calls. Given that our focus is on how firms use voluntary disclosure to impact the costliness of proposed transactions, we exclude any transaction that is already completed at the time of announcement—i.e., where the announcement is on or after the completion date (requirement 6).

The selection process, which we summarize in Table 1, results in a sample of 1,228 transactions. We report descriptive statistics for the full sample as well as for the conference call and non-conference call subsamples in Table 2. Merger-related conference calls occur for about 62 percent (758/1,228) of the sample transactions. A number of notable and statistically significant differences between conference call and non-conference call transactions are apparent. Consistent with prior research, bidders that hold conference calls are substantially larger, have smaller book-to-market ratios (*BM*), are more heavily represented in high-technology industries, and have greater analyst following and institutional ownership than firms that do not. The deal ratio is also larger for mergers involving conference calls, suggesting that bidders are more likely to engage in conference calls as the economic importance of the transaction increases. Bidders that hold conference calls finance a larger proportion of their mergers with stock, providing a preliminary

---

scheduled conference call to discuss the announced results, giving investors both the time and date of the conference call, and instructions on how to access the call; and Third, hold the conference call in an open manner, permitting investors to listen in either by telephonic means or through Internet webcasting" (SEC 2000, paragraph 4). The SEC has explicitly made road show presentations subject to Reg-FD requirements.

TABLE 1

Sample Selection Procedures

	Number of Bids
Total Merger or Acquisition transactions listed on the SDC database involving public bidders with non-missing data on transaction values, method of payment, and merger announcement dates and that were announced from January 1, 2002–December 31, 2006	12,840
Bidders without required variables on the CRSP/Compustat merged database	(4,602)
Transactions with a deal ratio of less than 10%	(6,186)
Duplicate transactions	(76)
Transactions where press releases could not be located in Factiva or on the bidding company's website	(539)
Announcements of already completed transactions	(209)
Final Sample	1,228

indication that disclosure incentives increase with the amount of the bidder's stock used to finance the transactions.[8]

Table 3 documents correlations. The correlations between conference calls and the remaining variables are consistent with the significant differences between the conference call and non-conference call subsamples highlighted in Table 2.

IV. DETERMINANTS OF MERGER-RELATED CONFERENCE CALLS

We analyze the determinants of firms' choice to hold conference calls at the merger announcement, using the following probit model:

$$\begin{aligned} CCALL_i = {} & \alpha_0 + \alpha_1 PCTSTOCK_i + \alpha_2 DEALRATIO_i + \alpha_3 INDR_i + \alpha_4 PRIV_i + \alpha_5 FOREIGN_i \\ & + \alpha_6 EARNANN_i + \alpha_7 LOGSIZE_i + \alpha_8 BM_i + \alpha_9 LOGANALYSTS + \alpha_{10} INSTOWN \\ & + \alpha_{11} REGULATED_i + \alpha_{12} HITECH_i + \alpha_{13} FINANCIAL_i + \text{yearly fixed effects} + \varepsilon_i \end{aligned} \tag{1}$$

where:

$CCALL$ = 1 for bidders that hold merger-related conference calls on the day of or the day after the merger announcement, and 0 for other bidders;
$PCTSTOCK$ = the value of stock consideration as a proportion of total consideration;
$DEALRATIO$ = the ratio of the total transaction value to the bidder's pre-announcement market value as of the beginning of the fiscal year;
$INDR$ = 1 when the bidder and the target have the same two-digit SIC code, and 0 otherwise;
$PRIV$ = 1 when the target firm is a private company, and 0 otherwise;
$FOREIGN$ = 1 when the target firm is a foreign company, and 0 otherwise;

[8] The 25th (75th) percentile of *PCTSTOCK* for the full sample is 0(67) percent. For the conference call subsample, the 25th (75th) percentile of *PCTSTOCK* is 0(50) percent, while for the conference call subsample, the 25th (75th) percentile of *PCTSTOCK* is 0(67) percent.

TABLE 2
Descriptive Statistics

Variable	Full Sample (n = 1,228) Mean	Full Sample (n = 1,228) Median	Conference Call (CCALL = 1) (n = 758) Mean	Conference Call (CCALL = 1) (n = 758) Median	Non-Conference Call (CCALL = 0) (n = 470) Mean	Non-Conference Call (CCALL = 0) (n = 470) Median	p-value for Mean Difference	p-value for Median Difference
MKTCAP	2,083	443	2,979	723	636	224	<0.001	<0.001
PCTSTOCK	0.322	0.000	0.363	0.116	0.256	0.000	<0.001	<0.001
DEALRATIO	0.590	0.291	0.632	0.350	0.523	0.229	0.029	<0.001
BM	0.651	0.526	0.577	0.493	0.770	0.589	<0.001	<0.001
INDR	0.366	0.000	0.346	0.000	0.398	0.000	0.065	0.065
PRIV	0.346	0.000	0.299	0.000	0.421	0.000	<0.001	<0.001
FOREIGN	0.108	0.000	0.120	0.000	0.087	0.000	0.071	0.071
EARNANN	0.090	0.000	0.124	0.000	0.034	0.000	<0.001	<0.001
REGULATED	0.048	0.000	0.051	0.000	0.043	0.000	0.479	0.479
HITECH	0.352	0.000	0.392	0.000	0.287	0.000	<0.001	<0.001
FINANCIAL	0.232	0.000	0.191	0.000	0.298	0.000	<0.001	<0.001
ANALYSTS	5.343	4.000	6.776	5.000	3.031	2.000	<0.001	<0.001
INSTOWN	0.542	0.589	0.616	0.651	0.424	0.386	<0.001	<0.001

The p-values are two-tailed. The tests of mean differences are based on the t-statistic, assuming unequal variances, and the tests for median differences are based on the Wilcoxon rank sums statistic.

Variable Definitions:

CCALL = 1 for bidders that hold merger-related conference calls on the day of or the day after the merger announcement, and 0 for other bidders;
MKTCAP = market value of equity at the beginning of the fiscal year in which the merger is announced (expressed in millions);
PCTSTOCK = value of stock consideration as a proportion of total consideration;
DEALRATIO = ratio of the total transaction value to the bidder's pre-announcement market value as of the beginning of the fiscal year;
BM = ratio of the bidder's book value of equity to its market value of equity as of the beginning of the fiscal year in which the merger is announced;
INDR = 1 when the bidder and target have the same two-digit SIC code, and 0 otherwise;
PRIV = 1 when the target firm is a private company, and 0 otherwise;
FOREIGN = 1 when the target firm is a foreign company, and 0 otherwise;
EARNANN = 1 when the merger announcement date is within five days of a quarterly earnings announcement date, and 0 otherwise;
REGULATED = 1 for firms with two-digit SIC codes of 48 or 49, and 0 otherwise;
HITECH = 1 for firms with two-digit SIC codes equal to 28, 35, 36, 73, or 87, and 0 otherwise;
FINANCIAL = 1 for firms with two-digit SIC codes in the range of 60-69, and 0 otherwise;
ANALYSTS = number of analysts issuing earnings forecasts for the fiscal quarter in which the merger is announced; and
INSTOWN = percentage of the acquirer's stock held by institutional investors as of the beginning of the quarter in which the merger is announced.

TABLE 3

Correlations

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
(1) *CCALL*		**0.35**	**0.14**	**0.16**	**−0.17**	**−0.05**	**−0.12**	**0.05**	**0.15**	0.02	**0.11**	**−0.12**	**0.37**	**0.28**
(2) *MKTCAP*	**0.15**		**−0.06**	**−0.21**	**−0.31**	−0.04	**−0.29**	0.03	−0.02	0.03	**−0.13**	**0.06**	**0.69**	**0.55**
(3) *PCTSTOCK*	**0.13**	**0.13**		**0.23**	**−0.09**	**−0.07**	**−0.07**	**−0.07**	−0.01	−0.02	**0.08**	**0.12**	0.04	**−0.18**
(4) *DEALRATIO*	**0.06**	−0.05	**0.19**		**0.15**	0.03	**−0.12**	−0.01	0.00	**0.05**	**−0.07**	**−0.06**	**−0.21**	**−0.09**
(5) *BM*	**−0.17**	**−0.10**	**−0.06**	**0.21**		0.00	−0.02	−0.04	−0.01	**0.07**	**−0.17**	**0.08**	**−0.24**	**−0.14**
(6) *INDR*	**−0.05**	0.02	**−0.07**	**0.06**	−0.04		0.03	−0.01	**0.06**	**−0.08**	−0.04	0.00	**−0.09**	0.00
(7) *PRIV*	**−0.12**	**−0.15**	**−0.10**	−0.03	0.00	0.03		**−0.05**	0.04	−0.02	**0.06**	−0.01	**−0.19**	**−0.11**
(8) *FOREIGN*	**0.05**	−0.02	**−0.06**	−0.04	−0.04	−0.01	**−0.05**		0.00	−0.02	**0.15**	**−0.16**	0.03	0.02
(9) *EARNANN*	**0.15**	0.02	−0.01	0.01	−0.03	**0.06**	0.04	0.00		0.01	**0.09**	**−0.05**	−0.01	0.03
(10) *REGULATED*	0.02	**0.07**	−0.02	**0.10**	**0.12**	**−0.08**	−0.02	−0.02	0.01		**−0.17**	**−0.12**	−0.01	−0.04
(11) *HITECH*	**0.11**	**−0.07**	**0.08**	−0.05	**−0.10**	−0.04	**0.06**	**0.15**	**0.09**	**−0.17**		**−0.41**	**0.06**	0.03
(12) *FINANCIAL*	**−0.12**	0.04	**0.12**	**−0.05**	−0.01	0.00	−0.01	**−0.16**	**−0.05**	**−0.12**	**−0.41**		**−0.12**	**−0.19**
(13) *ANALYSTS*	**0.33**	**0.40**	**0.12**	**−0.15**	**−0.19**	**−0.06**	**−0.17**	0.00	−0.01	0.00	**0.07**	**−0.10**		**0.56**
(14) *INSTOWN*	**0.29**	**0.11**	**−0.19**	**−0.14**	**−0.12**	0.00	**−0.11**	0.03	0.03	−0.04	0.04	**−0.19**	**0.46**	

Spearman (Pearson) correlation coefficients are presented above (below) the diagonal. The coefficients in bold are all statistically significant at less than the 10 percent level in two-tailed tests.

See Table 2 for variable definitions.

EARNANN = 1 when the merger announcement date is within five days of a quarterly earnings announcement date, and 0 otherwise;
LOGSIZE = the log of the market value of equity at the beginning of the fiscal year in which the merger is announced;
BM = the ratio of the bidder's book value of equity to its market value of equity as of the beginning of the fiscal year in which the merger is announced;
LOGANALYSTS = the log of the number of analysts issuing earnings forecasts for the fiscal quarter in which the merger is announced;
INSTOWN = the percentage of the acquirer's stock held by institutional investors as of the beginning of the quarter in which the merger is announced;
REGULATED = 1 for firms with two-digit SIC codes of 48 or 49, and 0 otherwise;
HITECH = 1 for firms with two-digit SIC codes of 28, 35, 36, 73, or 87, and 0 otherwise; and
FINANCIAL = 1 for firms with two-digit SIC codes in the range of 60–69, and 0 otherwise.

Although firms can engage in disclosure to influence stock price anytime between the announcement of the merger and its completion, we use the narrow event window consisting of day 0 and day 1 relative to the merger announcement when determining the value of *CCALL*. Our use of a narrow event window is based on the conventional wisdom that it is difficult to reverse initially negative market reactions to merger announcements. As Lipin and Sirower (2003, 23) state, "many companies have discovered [that] it's hard to put the genie back in the bottle once a deal gets a bad reception." Therefore, we expect bidders' incentives to engage in disclosure to favorably affect the market's response to a merger to be most pronounced at the time of the initial merger announcement.

We determine whether an acquirer held a conference call on the day of or the day after the merger announcement (i.e., whether *CCALL* is equal to 0 or 1) based first on whether the merger announcement press release indicated plans to hold a conference call during this interval. In cases where the merger announcement press release did not contain a conference call announcement, we searched all press releases issued by the acquirer for up to 14 days following the merger announcement for evidence of a separate conference call announcement.[9]

Our primary interest is on the effect of the economic significance of the deal (*DEALRATIO*) and of the method of payment (*PCTSTOCK*). However, we control for other merger characteristics that are likely to influence the degree of information asymmetry associated with the merger and the resulting intensity of investor demand for supplemental information. These factors include the degree of industry relatedness (*INDR*) between the merging entities, the target's foreign status (*FOREIGN*), and the target's private status (*PRIV*).

We also control for previously documented determinants of conference call use. Specifically, we control for *EARNANN* because it is common for U.S. companies to hold conference calls in connection with earnings announcements. Prior research has documented a positive association between firm size and the extent of voluntary disclosure as a result of greater investor demand for disclosure and greater economies of scale in disclosure (Kasznik and Lev 1995). Past conference call research also documents that the decision to hold a conference call is decreasing in *BM* (Frankel et al. 1999). Analysts and institutional investors are likely to require more information

[9] We identify 12 transactions where the acquirers held merger-related conference calls outside of our narrow merger announcement window. We set *CCALL* equal to 0 for these transactions because we are interested in the effect of the calls made at the time of the merger announcement on the initial market reaction to the merger announcement. Conference calls held after the merger announcement cannot affect the initial market reaction. However, none of our inferences change if we instead set *CCALL* equal to 1 for these 12 transactions.

from bidders and, because of Reg-FD, bidders can no longer provide information in private to them. Therefore, bidders' decisions to hold conference calls are likely to be influenced by analyst coverage and institutional ownership. Because the propensity to engage in a conference call likely varies across industries, we control for industry effects by incorporating separate indicator variables corresponding to the acquirer's membership in either high-tech, regulated, or financial industry categories. Finally, because companies' use of conference calls has increased dramatically over time, we control for year effects by including separate indicator variables corresponding to each of the announcement years represented in the sample.

Table 4 presents the results of the probit estimation of Equation (1). Conference calls are more likely for large acquirers and high-technology firms. The likelihood of a merger-related conference call also increases with analyst coverage and institutional holdings and decreases with book-to-market. More importantly, we find that bidders are more likely to hold conference calls in connection with the announcement of proposed mergers as the relative size of the deal (*DEALRATIO*) increases ($p < 0.001$) and as the proportion of consideration to be paid in the form of stock (*PCTSTOCK*) increases ($p < 0.001$).[10] The latter finding is consistent with our conjecture that stock-for-stock bidders are likely to respond to their greater need to favorably influence investor reactions by providing expanded disclosure in the form of conference calls at the time of the merger announcement.[11]

A potential alternative explanation for this finding is that stock is often used in large transactions, which likely require more communication. Our analysis controls for acquirer size (*LOGSIZE*) and the relative size of the transactions (*DEALRATIO*). To further control for the potential effect of transaction size, we sort the value of the transactions into quartiles and compare the rate of conference call use for stock transactions (i.e., those transactions where *PCTSTOCK* is 50 percent or greater) and cash transactions within each quartile. Untabulated results show that the proportion of stock payment increases with the size of the deal. However, within deals of similar size (i.e., within each quartile), the rate of conference call use is always significantly higher in stock deals than in cash deals.

V. THE EFFECT OF CONFERENCE CALLS ON THE MARKET REACTION TO MERGER ANNOUNCEMENTS

We next analyze whether the use of conference calls results in a more favorable investor reaction to merger announcements. To the extent that firms anticipate the market reaction to a

---

[10] The mean and standard deviation of *PCTSTOCK* are 32.2 percent and 40.0 percent, respectively. Therefore, the results in Table 3 indicate that a one standard deviation increase of 40.0 percent in *PCTSTOCK* from the mean of 32.2 percent to 72.2 percent is associated with an increase of approximately 8.71 percent in the likelihood of a merger-related conference call. Because a *PCTSTOCK* of 0 is 0.805 standard deviations below the mean *PCTSTOCK* of 32.2 percent and the unconditional probability of a merger-related conference call is 61.7 percent (758 conference call observations/1,228 total observations), the conditional probability of a merger-related conference call for observations where *PCTSTOCK* is equal to 0 is approximately 54.7 percent ($0.617 - [0.0871 \times 0.805]$). Similarly, because a *PCTSTOCK* of 1 is approximately 1.695 standard deviations above the mean *PCTSTOCK* of 32.2 percent and the unconditional probability of a merger-related conference call is 61.7 percent, the conditional probability of a merger-related conference call for observations where *PCTSTOCK* is equal to 1 is approximately 76.5 percent ($0.617 + [0.0871 \times 1.695]$). Hence, bidders that seek to finance their business combinations entirely with stock are 39.8 percent ($[76.5/54.7] - 1$) more likely to hold a merger-related conference call than bidders that intend to finance their transactions entirely with cash.

[11] Equation (1) treats *PCTSTOCK* as exogenous. This specification is reasonable under the assumption that bidders arrive at the deal details based on the economics of the transaction, independent of the conference call decision. There is a possibility, however, that the conference call decision is a determinant of the method of payment. To address this potential reverse causation, we employ an instrumental variables approach where we model the percentage of stock used in the transaction as a function of exogenous determinants suggested by prior literature. We then use the predicted value from this model as an instrumental variable for *PCTSTOCK* in Equation (1). Untabulated results show the conclusion that the proportion of stock in a transaction has a positive effect on the use conference calls still holds. Specifically, the coefficient on the instrumental variable is significantly positive (p-value < 0.001).

TABLE 4

Determinants of Merger-Related Conference Calls: Probit Estimation

$$CCALL_i = \alpha_0 + \alpha_1 PCTSTOCK_i + \alpha_2 DEALRATIO_i + \alpha_3 INDR_i + \alpha_4 PRIV_i + \alpha_5 FOREIGN_i + \alpha_6 EARNANN_i + \alpha_7 LOGSIZE_i + \alpha_8 BM_i + \alpha_9 LOGNALYST + \alpha_{10} INSTOWN + \alpha_{11} REGULATED_i + \alpha_{12} HITECH_i + \alpha_{13} FINANCIAL_i + \textit{yearly fixed effects} + \varepsilon_i$$

Dependent Variable: *CCALL*	Predicted Sign	Coefficient Estimate	p-value	Effect of Increase in Variable
Intercept	?	−2.008	<0.001	—
Method of Payment				
PCTSTOCK	+	0.603	<0.001	8.71%
Transaction Characteristics				
DEALRATIO	+	0.299	<0.001	9.20%
INDR	?	−0.103	0.228	−3.97%
PRIV	?	−0.002	0.977	−0.10%
FOREIGN	?	0.098	0.463	3.65%
Other Factors				
EARNANN	+	0.880	<0.001	25.84%
LOGSIZE	+	0.246	<0.001	15.17%
BM	−	−0.150	0.040	−3.13%
LOGANALYST	+	0.267	<0.001	8.90%
INSTOWN	+	0.322	0.031	3.82%
REGULATED	?	0.101	0.616	3.77%
HITECH	?	0.235	0.019	8.49%
FINANCIAL	?	−0.283	0.014	−11.07%
Yearly Fixed Effects	?	not tabulated	not tabulated	not tabulated
Pseudo R^2			22.15%	
n			1,228	

The p-values are one-tailed for coefficients with predicted signs and two-tailed otherwise.

Estimates are based on probit estimation. For continuous (binary) variables, the "Effect of Increase in Variable" is the change in the probability that a firm holds a conference call when the indicated variable is increased by one standard deviation (1) when all other continuous variables are set to their mean and all other binary variables are set to 0.

LOGSIZE is the log of the market value of equity at the beginning of the fiscal year in which the merger is announced. *LOGANALYST* is the log of the number of analysts issuing earnings forecast for the fiscal quarter in which the merger is announced.

See Table 2 for the definitions of the other variables.

merger announcement when deciding whether to hold a conference call, the relation between conference call use and the observed market reaction to the merger announcement is endogenous. This endogeneity may bias against finding a benefit to conference calls in an ordinary least-squares framework, particularly if acquirers are more motivated to hold conference calls to counteract anticipated unfavorable initial investor responses to their merger announcements. To account for the endogenous nature of the relation between conference call use and the observed market

reaction to merger announcements, we estimate the following model that incorporates the endogeneity correction set forth by Heckman (1979):[12]

$$
\begin{aligned}
CAR_i = {} & \beta_0 + \beta_1 CCALL_i + \beta_2 PCTSTOCK_i + \beta_3 PRIV_i + \beta_4 PCTSTOCK_i \times PRIV_i \\
& + \beta_5 DEALRATIO_i + \beta_6 INDR_i + \beta_7 FOREIGN_i + \beta_8 LOGSIZE_i + \beta_9 BM_i \\
& + \beta_{10} LOGANALYST_i + \beta_{11} INSTOWN_i + \beta_{12} LAMBDA_i \times CCALL_i + \beta_{13} LAMBDA_i \\
& \times (1 - CCALL_i) + \varepsilon_i \qquad (2)
\end{aligned}
$$

where *CAR* is the cumulative market-adjusted return over the three-day period spanning day 0 to day +2, where day 0 is the merger announcement date; and *LAMBDA* is the inverse Mills' ratio calculated according to Heckman (1979) based on the probit estimation of Equation (1).

The coefficient on *CCALL* (β_1) captures the reward to acquirers for engaging in conference calls. We include *PCTSTOCK* in the model based on prior evidence that bidders' announcement returns are generally negative for stock-for-stock acquisitions and positive for cash acquisitions.[13] We include *PRIV* and interact it with *PCTSTOCK* based on the finding in Chang (1998) that the private status of a target mitigates the negative market reaction to stock-for-stock mergers. We include *DEALRATIO* as a control for the economic importance of the merger. Prior studies document that this variable is positively related to bidders' announcement returns (Asquith et al. 1983). We include *INDR* because Morck et al. (1990) and Maquieira et al. (1997) find that returns to bidding firms are lower when the merger is diversifying. We include *FOREIGN* to capture the possibility that foreign targets' lack of visibility increases investor uncertainty and skepticism, thereby leading to lower announcement returns.[14] We include *LOGSIZE* because Loderer and Martin (1997) find that size is negatively correlated with the bidders' merger announcement stock returns. We control for *BM* based on evidence that Tobin's Q and book-to-market are negatively associated with bidder returns (e.g., Lang et al. 1989; Servaes 1991; Gong et al. 2008a). Moeller et al. (2007) suggest that acquirers' abnormal returns are negatively related to information asymmetry and diversity of opinion. We proxy for these factors by the acquirers' analyst coverage (*LOGANALYST*) and proportion of institutional holdings (*INSTOWN*). Following Tucker (2007), we allow the coefficient on the inverse Mills' ratio to differ for conference call and non-conference call transactions.[15]

We include the industry membership and announcement year indicator variables in the selection model (Equation (1)) and not in the outcome model (in Equation (2)) because these variables are highly associated with the conference call decision (they are jointly significant in Equation (1) at p = 0.001) but have no theorized effect on announcement returns. There is little doubt that industry membership and announcement year indicators are exogenous. We therefore use them as instrumental variables in the selection model. An untabulated test of over-identifying restrictions recommended by Larcker and Rusticus (2010) fails to reject the null hypothesis of exogeneity.

---

[12] We exclude merger announcements that coincide with earnings announcements in order to avoid the contaminating effects of other information.

[13] See, for example, Travlos (1987), Lang et al. (1989), Servaes (1991), Chang (1998), Fuller et al. (2002), Shleifer and Vishny (2003), Louis (2004), Moeller et al. (2005), Malmendier and Tate (2008).

[14] The home bias literature indicates that domestic investors lack visibility in foreign firms (e.g., Kang and Stulz 1997; Ahearne et al. 2004; Covrig et al. 2001; Suh 2001; Bradshaw et al. 2004).

[15] Our results are unchanged if we constrain the coefficient on the inverse Mills' ratio to be the same for conference call and non-conference call transactions.

Therefore, industry membership and announcement year indicators represent suitable instrumental variables for identifying the system of equations.

Table 5 presents the results on the effect of conference calls on the market reaction to merger announcements. The univariate results in Panel A show that the average merger announcement abnormal return is lower for firms that hold conference calls than for those that do not. This finding is not surprising given that the decision to hold a merger-related conference call is positively related to factors that are generally associated with negative merger announcement returns, stock financing in particular. That is, an acquirer is more likely to hold conference calls when the nature of the transaction is such that the market is likely to react negatively in the absence of supplementary disclosure.

The results in Table 5, Panel B show the association between conference calls and the market reactions to merger announcements, after controlling for the potential selection bias related to the decision to hold merger-related conference calls and other determinants of the market reaction. Many of the coefficients on the control variables are significant and their signs are largely consistent with expectations, lending validity to the model. Turning to the main result, bidders that provide conference calls concurrent with their merger announcements experience announcement period returns that are about 6.5 percent ($p = 0.016$) higher than they would otherwise experience.

Although our discussion in Section II explains why stock acquirers have greater motivations than cash acquirers to hold merger-related conference calls, it is not clear (and we offer no expectation about) whether the realized benefits to conference calls are similar for stock and cash acquirers. In an untabulated supplemental test, we address this question by augmenting Equation (2) with an interaction of *PCTSTOCK* and *CCALL*. Because *CCALL* is endogenous, an interaction involving *CCALL* is also endogenous, and simply including the interaction in the model without an appropriate correction could lead to inconsistent estimates (Wooldridge 2002). We, therefore, follow the procedure suggested by Wooldridge (2008) for interacting binary endogenous variables and exogenous regressors. The coefficient on the resulting interaction is insignificant, which indicates no difference in the benefit to conference calls for cash versus stock acquirers. This finding likely reflects the fact that, once managers have decided to hold a conference call, the benefit will not necessarily vary across cash and stock transactions since managers are likely to hold conference calls when the firm is most likely to benefit from the call, irrespective of the method of payment.

Given the evidence that bidders that provide conference calls concurrent with their merger announcements experience higher announcement period returns than they would otherwise experience, the question that naturally emerges is: "Why are all merger announcements not accompanied by conference calls?" While the majority (62 percent) of the transactions in the full sample is accompanied by conference calls, the rate of conference call use is substantially less than 100 percent. We offer several conjectures as to why this is the case. First, the decision to hold conference calls is the result of a cost-benefit analysis. The benefits are more likely to accrue to acquirers that would have been over-penalized by the market because of the complexities of their transactions, the signal associated with the nature of the transactions, or some other reasons. Although providing a conference call has the potential to increase investor confidence and induce a more favorable (or less adverse) market reaction, voluntary disclosure is also costly. Proprietary information might be revealed in the process. Forward-looking disclosures expose managers to reputation loss and potential litigation (Francis et al. 1994; Skinner 1997). In addition, conference calls are a risky form of disclosure because, if managers' responses to analyst inquiries during conference calls do not pass muster, investor confidence could be undermined rather than

TABLE 5

The Effect of Merger-Related Conference Calls on the Market Reaction to Merger Announcements

Panel A: Univariate Comparisons of Event-Period Returns (p-values in parentheses)

	Conference Call (n = 653)	Non-Conference Call (n = 471)	Difference
Mean	0.005	0.023	−0.018
	(0.181)	(<0.001)	(0.001)
Median	−0.003	0.008	−0.011
	(0.710)	(<0.001)	(0.009)

Panel B: Multivariate Analysis of Event-Period Returns

$$CAR_i = \beta_0 + \beta_1 CCALL_i + \beta_2 PCTSTOCK_i + \beta_3 PRIV_i + \beta_4 PCTSTOCK_i \times PRIV_i + \beta_5 DEALRATIO_i + \beta_6 INDR_i + \beta_7 FOREIGN_i + \beta_8 LOGSIZE_i + \beta_9 BM_i + \beta_{10} LOGANALYST_i + \beta_{11} INSTOWN_i + \beta_{12} LAMBDA_i \times CCALL_i + \beta_{13} LAMBDA_i \times (1 - CCALL_i) + \varepsilon_i.$$

Variables	Predicted Sign	Coefficient Estimate	p-value
Intercept	?	0.063	<0.001
CCALL	+	0.065	0.016
PCTSTOCK	−	−0.070	<0.001
PRIV	?	−0.006	0.374
PCTSTOCK × *PRIV*	+	0.077	<0.001
DEALRATIO	+	0.008	0.051
INDR	−	0.010	0.964
FOREIGN	−	−0.009	0.148
LOGSIZE	−	−0.009	0.001
BM	−	0.002	0.685
LOGANALYST	−	−0.011	0.022
INSTOWN	−	−0.029	0.008
LAMBDA × *CCALL*	?	−0.031	0.153
LAMBDA × (1 − *CCALL*)	?	−0.045	0.013
n		1,099	
Adjusted R^2		14.44%	

Panel C: Multivariate Analysis of Long-Run Returns

$$LTRETURNS_i = \beta_0 + \beta_1 CCALL_i + \beta_2 PCTSTOCK_i + \beta_3 PRIV_i + \beta_4 PCTSTOCK_i \times PRIV_i + \beta_5 DEALRATIO_i + \beta_6 INDR_i + \beta_7 FOREIGN_i + \beta_8 LOGSIZE_i + \beta_9 BM_i + \beta_{10} LOGANALYST_i + \beta_{11} INSTOWN_i + \beta_{12} LAMBDA_i \times CCALL_i + \beta_{13} LAMBDA_i \times (1 - CCALL_i) + \varepsilon_i.$$

Variables	Coefficient Estimate	p-value
Intercept	−0.084	0.338
CCALL	−0.136	0.419
PCTSTOCK	−0.037	0.466

(continued on next page)

Panel C: Multivariate Analysis of Long-Run Returns

$$LTRETURNS_i = \beta_0 + \beta_1 CCALL_i + \beta_2 PCTSTOCK_i + \beta_3 PRIV_i + \beta_4 PCTSTOCK_i \times PRIV_i + \beta_5 DEALRATIO_i + \beta_6 INDR_i + \beta_7 FOREIGN_i + \beta_8 LOGSIZE_i + \beta_9 BM_i + \beta_{10} LOGANALYST_i + \beta_{11} INSTOWN_i + \beta_{12} LAMBDA_i \times CCALL_i + \beta_{13} LAMBDA_i \times (1 - CCALL_i) + \varepsilon_i.$$

Variables	Coefficient Estimate	p-value
PRIV	−0.008	0.830
PCTSTOCK × *PRIV*	−0.012	0.879
DEALRATIO	0.029	0.320
INDR	−0.036	0.233
FOREIGN	−0.022	0.649
LOGSIZE	0.032	0.035
BM	0.027	0.422
LOGANALYST	−0.009	0.977
INSTOWN	−0.025	0.703
LAMBDA × *CCALL*	0.008	0.947
LAMBDA × (1 − *CCALL*)	0.133	0.189
n	1,045	
Adjusted R^2	1.43%	

The p-values are one-tailed for coefficients with predicted signs and two-tailed otherwise. In Panel A, the tests of mean differences are based on the t-statistic, assuming unequal variances, and the tests for the median differences are based on the Wilcoxon rank sums statistic.

CAR is the cumulative market-adjusted return over the three-day period spanning day 0 to day +2, where day 0 is the merger announcement date. *LTRETURNS* is one-year buy-and-hold abnormal return starting on the third day after the merger announcement using the match-firm approach suggested by Barber and Lyon (1997). *LAMBDA* is the inverse Mills' ratio calculated according to Heckman (1979) based on the probit estimation of Equation (1) in Table 4.

See Table 2 for remaining variable definitions.

enhanced.[16]

Second, conference calls are likely to be beneficial only to those acquirers that have favorable information to impart, in addition to the information in the press releases and the observable characteristics of the mergers. Hutton et al. (2003) provide evidence that the market reacts to good news disclosures only when they are augmented with sufficient detail to make the good news credible. Otherwise, the market views good news assertions by managers as "cheap talk." Hence, it appears that, to benefit from conference calls in the form of a more favorable market reaction, firms must have good news that they can support with sufficient detail to make the good news

[16] For example, analysts reacted skeptically to AT&T's initial conference call in connection with its proposed purchase in 1998 of TCI. Of particular concern to analysts was the plausibility of AT&T's growth projections. The concern expressed by analysts was so great that AT&T felt compelled to hold a second conference call only days after its initial conference call. In the second call, AT&T discussed the technical details and anticipated costs of integrating AT&T's and TCI's technologies. However, analysts responded to the second conference call by continuing to question many aspects of the deal, particularly AT&T's assumptions. This case provides an example of a situation where providing additional details actually fuels rather than alleviates analyst and, by extension, investor skepticism. Highlighting this risk, we document in untabulated analysis of a matched sample that investor response to the announcement of stock-for-stock transactions is more favorable for the conference call transactions 62 percent of the time. While this represents a clear majority of the transactions, the 38 percent of the time where conference call transactions are greeted with less favorable reactions is nontrivial.

credible. Therefore, it seems reasonable to assume that bidders hold conference calls when they believe that they have favorable information that they can impart through this disclosure medium. In Section VII, we provide evidence on the type of information provided during conference calls that appears to drive the more favorable reaction.

VI. DO ACQUIRERS USE CONFERENCE CALLS TO HYPE THEIR STOCK PRICES?

Acquirers that provide conference calls concurrent with their merger announcements could experience superior market performance at merger announcements because managers use the conference calls to hype their stock prices. We test this conjecture by analyzing the acquirers' post-merger announcement abnormal returns and pre-merger announcement abnormal accruals.

Post-Merger Announcement Abnormal Returns

If managers use conference calls to temporarily inflate their stock prices prior to the transactions, then the relatively favorable market reaction we document would be temporary. Therefore, we compare the acquirers' abnormal returns over the year after the merger announcement across acquirers that hold conference calls and those that do not. If the managers use the conference calls around merger announcements to temporarily hype their stock prices, then we expect the post-merger announcement abnormal returns to be more negative when the merger announcements are accompanied by conference calls.

We estimate a firm's long-term abnormal returns (*LTRETURN*) by adjusting its buy-and-hold return by the return of a matched firm, as suggested by Barber and Lyon (1997) and Kothari and Warner (1997). Following Barber and Lyon (1997), we choose the match firm from all firms with a market value of equity between 70 percent and 130 percent of the market value of equity of the acquirer as of the beginning of the quarter in which the merger was announced. From this set of firms, we match the acquirer with the firm that has the closest book-to-market ratio.

We then reestimate Equation (2) after replacing *CAR* with *LTRETURN*. The results, reported in Table 5, Panel C provide no evidence that firms that hold conference calls concurrent with their merger announcements have worse post-merger announcement market performance than those that do not hold conference calls (i.e., β_1 is insignificant). Therefore, there is no evidence that the more favorable market reaction to merger announcements that are accompanied with conference calls subsequently reverses, which suggests that, in general, managers use conference calls around merger announcements as a means to convey genuine private information as opposed to falsely hyping their stock prices.

Pre-Merger Announcement Abnormal Accruals

Prior studies also suggest that stock-for-stock acquirers tend to report higher pre-merger announcement abnormal accruals (Erickson and Wang 1999; Louis 2004). Louis (2004) also suggests that investors are generally misled by the pre-merger announcement abnormal accruals. Therefore, to further examine whether acquirers that provide conference calls concurrent with their merger announcements are motivated by the desire to hype their stock prices, we analyze the association between stock-for-stock acquirers' use of conference calls and the abnormal accruals they include in the earnings reported within 90 days prior to the merger announcement.[17] If

[17] We classify transactions as stock-for-stock if 50 percent or more of the consideration is comprised of the acquirer's stock. We use a 50 percent threshold to isolate transactions where the incentives associated with the use of stock are pronounced, while ensuring that we have enough observations to conduct our analysis. The average percentage of stock financing for our stock-for-stock subsample using this threshold is about 85 percent.

acquirers that use conference calls around merger announcements are motivated by the desire to hype their stock prices, then these acquirers would presumably have a stronger motivation to report higher pre-merger announcement abnormal accruals.

We measure abnormal accruals by the residual from a modified Jones (1991) model similar to the model in Louis and White (2007) and Gong et al. (2008a, 2008b). Specifically, for each calendar quarter and two-digit SIC-code industry, we estimate the following model using all firms that have the necessary data on Compustat:

$$TA_i = \sum_{j=1}^{4} \lambda_{j-1} Q_{j,i} + \lambda_4(\Delta SALE_i - \Delta AR_i) + \lambda_5 PPE_i + \lambda_6 LTA_i + \varepsilon_i \tag{3}$$

where:

TA = total accruals;
Q_j = a binary variable taking the value of 1 for *fiscal* quarter j, and 0 otherwise;
$\Delta SALE$ = the quarterly change in sales;
ΔAR = quarterly change in accounts receivable;
PPE = property, plant, and equipment at the beginning of the quarter; and
LTA = the fourth lag of total accruals.

We measure total accruals as earnings before extraordinary items and discontinued operations (Compustat quarterly data item #76) minus operating cash flows (#108 – #78). If a firm does not have operating cash flows on Compustat, then we compute total accruals as change in non-cash current assets (change in item #40 minus change in item #36) minus change in current liabilities (change in data item #49) plus change in debt in current liabilities (change in data item #45) minus depreciation (data item #5). All variables are scaled by total assets at the beginning of the quarter. To mitigate the effect of outliers and errors in the data, for each calendar quarter, we delete the top and bottom one-percentiles of the deflated TA, ($\Delta SALE - \Delta AR$), and PPE. We then delete the top and bottom one-percentiles of the regression residual (our abnormal accrual estimate). We also require at least 10 observations for each estimation.

Following Kothari et al. (2005), we adjust the estimated abnormal accruals for performance. For each quarter and each industry (two-digit SIC code), we create five portfolios by sorting the data into quintiles based on the return-on-assets from the same quarter in the previous year.[18] The performance-matched abnormal accruals for a sample firm are the firm-specific abnormal accruals minus the mean (excluding the sample firm) abnormal accruals for its respective industry-performance-matched portfolio.

Table 6 reports the abnormal accruals for the quarter prior to the merger announcement. Consistent with prior studies, the abnormal accruals are positive; however, they are substantially larger for those acquirers that do not hold conference calls at the time of the merger announcement. The difference in abnormal accruals between stock-for-stock acquirers that hold conference calls and those that do not is 0.0158 (two-tailed p-value of 0.047). One plausible explanation for this result is that stock-for-stock acquirers that inflate their accruals are more likely to refrain from holding conference calls to avoid questions about their financial reports. To provide evidence on this possibility, we also report differences in the pre-announcement accruals across high and low

[18] Ideally, we would match on the current quarter's ROA. However, this is not practical because current ROA (i.e., the ROA for the quarter immediately before the merger announcement) is affected by the pre-merger announcement earnings management that we are trying to capture. Our matching procedure is consistent with Louis (2004), Fischer and Louis (2008), Louis and White (2007), Gong et al. (2008a, 2008b), and Louis et al. (2008), among others.

TABLE 6

Stock-for-Stock Acquirers' Average Abnormal Accruals Conditional on Conference Calls and the Level of Analyst Coverage

	(1) Non-Conference Call		(2) Conference Call		(2)−(1) Difference	
	Mean	**Median**	**Mean**	**Median**	**Mean**	**Median**
All	0.0218	0.0149	0.0060	0.0077	−0.0158	−0.0072
	(0.007)	(0.008)	(0.058)	(0.027)	(0.047)	(0.047)
	{n = 53}	{n = 53}	{n = 187}	{n = 187}		
Low analyst coverage	0.0249	0.0125	0.0063	0.0085	−0.0186	−0.0040
	(0.009)	(0.016)	(0.116)	(0.070)	(0.052)	(0.072)
	{n = 35}	{n = 35}	{n = 102}	{n = 102}		
High analyst coverage	0.0157	0.0161	0.0057	0.0050	−0.0100	−0.0110
	(0.169)	(0.123)	(0.155)	(0.155)	(0.280)	(0.188)
	{n = 18}	{n = 18}	{n = 85}	{n = 85}		

Two-tailed p-values are reported in parentheses. The tests of mean differences are based on the t-statistic, assuming unequal variances, and the tests for median differences are based on the Wilcoxon rank sums statistic.

The abnormal accruals are for the quarter immediately preceding the merger announcement by stock-for-stock acquirers. Accruals are scaled by lagged total assets. The measurement process is described in Section VI.

analyst followings. We find no evidence that the lower magnitude of accruals for conference call transactions is more pronounced for the high analyst coverage subsample, which would be expected if acquirers that engage in accrual inflation seek to avoid the analyst scrutiny associated with conference calls. Thus, the results seem more consistent with the notion that acquirers that hold conference calls concurrent with stock issuances are motivated more by a desire to convey genuine private information than to falsely hype their stock prices.

VII. SMALL SAMPLE EVIDENCE ON THE INFORMATION CONTENT OF CONFERENCE CALLS

In Section V, we provide evidence that merger announcements accompanied by conference calls experience relatively more favorable market reactions and suggest that this finding is due to the information provided by management during these calls. In this section, we use content analysis on a subsample of transactions in order to provide direct evidence on the information content of conference calls. We then examine whether the information content of the conference calls contributes to the relatively more favorable response to merger announcements that are accompanied by conference calls. Because of the cost of coding conference call disclosures, we perform this content analysis on a subsample of transactions.

Identification and Characteristics of the Subsample

The subsample for the content analysis consists of those transactions where conference call transcripts were available either from FD Wires or StreetEvents and where a suitable control transaction could be located. The objective of our matching process is to identify transactions that were as economically similar as possible to the treatment transactions to maximize the ability to attribute differences in the information disclosed by the conference and non-conference call acquirers to managerial choice rather than to differences in the economics of the mergers.

We impose an initial requirement that the estimated likelihood of conference call use from estimating Equation (1) for the control transaction be within 20 percent of the corresponding conference call transaction.[19] For the subset of potential control transactions meeting this criterion, we impose the following additional criteria designed to make sure the final pairs have similar transaction characteristics: (1) the percentage of stock for the control transaction is within 5 percent of the percentage of stock for the conference call transaction; (2) the difference in deal ratio between the treatment and control transactions is less than 20 percent; (3) the percentage difference in acquirer size between the treatment and control transactions is less than 20 percent; and (4) the transaction takes place within a year of the treatment transaction.

In cases where we identify multiple control transactions meeting the above criteria for a given treatment transaction, we retain the control transaction that is nearest to the treatment transaction on the basis of the estimated probability of conference call use. This process leads to the identification of 147 treatment-control pairs of cash-financed mergers and 20 treatment-control pairs of stock-for-stock mergers.[20] Table 7 provides comparative descriptive statistics on the conference call transactions and their corresponding control transactions. The differences between the characteristics of the conference call and control transactions are insignificant, with the exception of the *DEALRATIO* in the case of the cash transactions. Thus, we believe that our matching procedures are successful in identifying control transactions that are economically similar to the conference call transactions.

Table 7 also compares *CAR* across the treatment and the control firms. The results reported in Table 7, Panel A provide no evidence that investors react more favorably to the announcement of a cash merger for the conference call group than for the non-conference call group. The mean (median) pairwise difference in *CAR* is −0.66 percent (−0.67 percent), with a one-tailed p-value of 0.843 (0.842). In contrast, the results reported in Panel B of Table 7 for the stock-for-stock mergers show that, on average, *CAR* is higher for the conference call transactions than for the control transactions. The mean (median) pairwise difference in *CAR* between conference call and non-conference call firms of 4.79 percent (3.90 percent) is significant with a one-tailed p-value of 0.014 (0.016), which indicates that investors react more favorably to the announcement of a stock-for-stock merger when the announcement is accompanied by a conference call.[21] Our failure to find evidence of benefits to conference calls for cash acquirers in the matched sample contrasts with the substantial benefits we document in Table 5 for the full sample and is likely attributable to differences between the matched sample and the full sample.[22] Since evidence of relatively

---

[19] The estimated probability of conference call use is a convenient matching criterion because it parsimoniously captures the factors relevant to the managers' conference call decisions. However, using it as the only match criterion results in significant pairwise differences on dimensions directly relevant to the economics of the transactions, such as: percentage of stock, deal ratio, and acquirer size. Therefore, we use the estimated probability as the first but not the only matching criterion.

[20] We obtain few matched pairs for the stock transactions because of the lack of mergers with similar acquirer size and deal ratio that were announced within a year of the announcement date for the conference call transactions.

[21] To ensure that our results are not due to a failure to adequately control for differences between the conference call and control transactions, we also regress the pairwise differences in *CAR* against pairwise differences in size, percentage of stock, deal ratio, book-to-market, analyst following, and the estimated probability of conference call use based on Equation (1). The conditional mean pairwise difference in *CAR* between conference-call and non-conference-call acquirers is captured by the regression intercept. The results from this multivariate analysis are consistent with our univariate analysis. Specifically, the conditional mean pairwise difference in *CAR* between the two groups of stock-for-stock acquirers is 5.03 percent (one-tailed p-value = 0.01). In contrast, for cash transactions, the mean pairwise difference is −1.23 percent (two-tailed p-value of 0.09).

[22] It is worth noting that the subsample differs from the full sample in terms of market capitalization (average market capitalization for this subsample of $625 million versus the full sample average of $2 billion). The average deal ratio from the subsample of 28 percent is also smaller than the average for the full sample of 59 percent. These differences reflect the difficulty in identifying matches for the largest firms and transactions in our sample for which comparable

more favorable market responses to conference call transactions is limited to stock transactions, we limit our detailed content analysis to those transactions.

Development of Taxonomy Used to Classify Merger-Related Disclosures

At the outset of the project, we initially read and transcribed all disclosures made in a small random selection of press releases and conference call transcripts. We then grouped the disclosures into similar categories based on the nature of the disclosure, which required judgment. Given this subjectivity of the classification process, we include an Appendix to provide transparency on our classification choices.

As indicated in the Appendix, we classify the disclosures from the initial transcription process into six major categories: (1) deal details, which outline the terms of the transaction; (2) acquirer business, which provides historical information on the acquirer's business; (3) target business, which provides historical information on the target's business; (4) strategic rationale, which outlines the strategic rationale for the merger; (5) integration details, which outline plans for post-merger consolidation of operations; and (6) financial guidance, which provides guidance on the financial impact of the merger. We characterize information from the first three categories as historical information and information from the last three categories as forward-looking information.

Coding Procedures

We applied the classification scheme set forth in the Appendix to the press releases and transcripts that comprise our matched sample. The process involved a line-by-line reading of the press release and/or conference call transcripts. For each sentence or related group of sentences, we determined the nature of the disclosure made, the number of words devoted to the disclosure, and referred to the classification scheme in the Appendix to determine the appropriate classification of the disclosure.[23]

When counting the number of words in the conference call transcripts, we ignored redundancies between the conference call and the press release disclosures because we are interested in the incremental information provided in the conference calls.[24] In addition, because we were only interested in the nature and amount of information provided by management, we included only management disclosures in the word count and exclude words attributable to analysts during the question-and-answer (Q&A) session of the conference calls.

Within-Transaction Comparisons

For the 20 stock transactions that are accompanied by conference calls, we examine the information content of the press releases and the accompanying conference calls. We measure the

---

firms do not exist. Despite the differences, the subsample does exhibit the finding in the full sample that the conference call transactions have relatively more favorable reactions than the non-conference call transactions. Therefore, content analysis for this subsample is likely to provide insights on the drivers of the relatively more favorable reaction to conference call transactions that we document for the full sample.

23 The coding process was admittedly subjective. However, in the course of coding, we encountered no disclosures that were not contained in the initial coding scheme set forth in the Appendix. Therefore, once the coding scheme was agreed upon, we had to exercise very little judgment about how disclosures should be classified. As a further check on the consistency of our coding, we hired a research assistant to independently code the press releases and conference call transcripts based on the coding scheme we provided. This research assistant was unaware of our expectations and had no exposure to any of the data. We found that our inferences were unchanged using the research assistant's coding.

24 Redundancies are primarily from the following sources. In the prepared remarks section of the conference call, management's "script" often includes direct quotations from the press release. Even in cases where management does not recite direct passages from the press release, they often repeat basic details that are already contained in the press release such as the amount or form of consideration. There are also instances during the question-and-answer portion of the call when analysts requested that management repeat details disclosed earlier in the call. In all such cases, the coding author excluded the repeated disclosures when coding the information content of the conference call.

TABLE 7

Transaction and Bidder Attributes for a Sample of Merger Announcements Accompanied by Conference Calls and a Control Sample of Merger Announcements with Press Releases Only

Panel A: Cash Transactions (n = 147)

	Conference Call Sample	Control Sample	Pairwise Difference Conference Call – Control	p-value of Pairwise Difference
	Mean [Median]	Mean [Median]	Mean [Median]	Mean [Median]
MKTCAP (millions)	871	897	−26	0.075
	[466]	[472]	[−2]	[0.351]
PCTSTOCK	0.00	0.00	0.00	0.159
	[0.00]	[0.00]	[0.00]	[0.500]
DEALRATIO	0.28	0.25	0.03	<0.001
	[0.23]	[0.18]	[0.02]	[<0.001]
BM	0.625	0.670	−0.045	0.319
	[0.572]	[0.532]	[−0.043]	[0.546]
ANALYSTS	5.05	4.80	0.26	0.391
	[5.00]	[4.00]	[0.00]	[0.180]
PROBABILITY	54.60	52.50	2.10	0.143
	[55.05]	[53.60]	[0.44]	[0.133]
CAR (%)	1.30	1.96	−0.66	0.843
	[0.62]	[1.34]	[−0.67]	[0.842]
LTRETURNS (%)	−5.11	0.52	−5.63	0.073
	[−3.04]	[−5.05]	[−2.25]	[0.240]

Panel B: Stock-for-Stock Transactions (n = 20)

	Conference Call Sample	Control Sample	Pairwise Difference Conference Call – Control	p-value of Pairwise Difference
	Mean [Median]	Mean [Median]	Mean [Median]	Mean [Median]
MKTCAP (millions)	626	623	3	0.863
	[415]	[473]	[−7]	[0.596]
PCTSTOCK	0.92	0.91	0.01	0.028
	[1.00]	[1.00]	[0.00]	[0.063]
DEALRATIO	0.28	0.28	0.00	0.844
	[0.21]	[0.20]	[0.01]	[0.841]
BM	0.560	0.605	−0.045	0.571
	[0.492]	[0.523]	[−0.011]	[0.784]
ANALYSTS	5.65	4.90	0.75	0.487
	[4.00]	[4.00]	[0.00]	[0.692]
PROBABILITY	66.24	62.81	3.43	0.143
	[63.08]	[66.12]	[2.41]	[0.133]
CAR (%)	2.17	−2.62	4.79	0.014
	[0.27]	[−3.64]	[3.90]	[0.016]
LTRETURNS (%)	10.72	−10.76	21.48	0.355
	[−9.42]	[−8.29]	[5.27]	[0.546]

(continued on next page)

All p-values are two-tailed, except those related to *CAR*, which are one-tailed. The significance of mean pairwise differences is based on one-sample t-tests and the significance of median pairwise differences is based on the Wilcoxon sign-rank test.

PROBABILITY is the estimated probability of holding a conference call based on the probit estimation of Equation (1).

The other variables are defined as in Tables 2, 4, and 5.

volume of information provided using word count. In untabulated analysis, we find that the conference call provides an average of 4.3 times more total volume of information than the press release ($p < 0.01$). Moreover, conference calls provide significantly greater volume of information across all information categories except for acquirer business and deal details. Table 8, Panel A summarizes within-transaction comparisons of the relative emphasis on information categories in merger announcement press releases versus conference call transcripts for the 20 treatment transactions. We calculate the relative emphasis devoted to a category as the proportion of the total word count attributable to that category. While only 45 percent of the discussion in press releases pertains to forward-looking details, over 70 percent of the disclosure volume in conference calls is devoted to forward-looking information.

We also examine differences in the information content of the prepared-remarks section of the conference call versus the Q&A session. A key feature of conference calls is the ability of call participants to probe management for additional details during a Q&A session that typically follows prepared remarks by management. Therefore, we are interested in how important these Q&A sessions are in eliciting additional information from management. Untabulated analysis indicates that the Q&A portion of conference calls has roughly equal volume to the prepared remarks portion (mean and median pairwise differences are insignificant). Thus, the opportunity conference calls offer to analysts to probe management appears to be a major source of the total information provided in these calls.[25]

Table 8, Panel B summarizes within-transaction comparisons of the relative emphasis on information categories in the prepared remarks portion of the conference call versus the Q&A portion. There is no significant difference in the overall emphasis placed on historical or forward-looking information in the Q&A session versus the press release. However, there is a greater emphasis on integration details during the Q&A portion, while there is a significantly greater discussion of the acquirer business and the strategic rationale for the transaction in the prepared remarks section.

Between-Transaction Comparisons

We next compare the total information provided by the treatment and control samples. The volume of information provided in the conference call transactions is the sum of the word counts contained in the press release and in the related conference call transcripts. We aggregate across these information sources because we are interested in comparing the overall volume of disclosure provided by conference call transactions to that provided in control transactions.[26] In untabulated

[25] In contrast to our findings, Matsumoto et al. (2006) find, in the context of earnings announcement conference calls, that a greater proportion of the total call volume occurs during the Q&A portion of the conference call. These contrasting findings highlight the different insights that are gained by examining conference calls outside of the earnings contexts that have been the primary focus of prior studies.

[26] Recall that in computing word count totals for conference call transcripts, we avoid including any conference call disclosures that are redundant with disclosures made in the corresponding press release. Thus, double counting of disclosures is not an issue when aggregating word counts across the two information sources.

TABLE 8

Information Provided in Conference Calls and Press Releases for a Subsample of 20 Stock-for-Stock Merger Announcements Where the Bidders Provide Both Press Releases and Conference Calls

Panel A: Press Releases versus Conference Call: Within-Transaction Comparisons

	(1) Press Release		(2) Conference Call		(1)−(2) Pairwise Difference	
	Mean	Median	Mean	Median	Mean (p-value)	Median (p-value)
Deal details	27.91	26.47	9.02	8.48	18.89 (<0.001)	16.55 (<0.001)
Acquirer business	14.78	11.85	4.02	2.77	10.76 (0.001)	10.52 (0.001)
Target business	12.00	10.91	14.42	11.59	−2.41 (0.351)	1.06 (0.891)
Strategic rationale	32.95	33.18	43.91	42.71	−10.96 (0.006)	−8.96 (0.008)
Integration details	8.03	7.05	18.76	16.27	−10.73 (<0.001)	−9.83 (<0.001)
Financial guidance	4.32	2.65	9.87	10.18	−5.55 (0.006)	−5.21 (0.011)
Total forward-looking details	45.31	43.85	72.55	77.71	−27.24 (<0.001)	−28.13 (<0.001)

Panel B: Prepared Remarks versus Q&A Portion: Within-Transaction Comparisons

	(1) Prepared Remarks		(2) Q&A		(2)−(1) Pairwise Difference	
	Mean	Median	Mean	Median	Mean (p-value)	Median (p-value)
Deal details	9.56	8.82	8.52	4.13	−1.04 (0.695)	−0.39 (0.622)

(continued on next page)

Panel B: Prepared Remarks versus Q&A Portion: Within-Transaction Comparisons

	(1) Prepared Remarks		(2) Q&A		(2)−(1) Pairwise Difference	
	Mean	**Median**	**Mean**	**Median**	**Mean (p-value)**	**Median (p-value)**
Acquirer business	6.26	3.85	1.48	0.00	−4.79	−3.24
					(<0.001)	(<0.001)
Target business	11.12	8.37	13.54	6.48	2.42	0.27
					(0.557)	(0.709)
Strategic rationale	52.36	49.15	38.92	39.21	−13.44	−12.39
					(0.029)	(0.019)
Integration details	12.86	11.12	24.71	20.85	11.85	10.89
					(0.005)	(0.007)
Financial guidance	7.83	6.52	12.83	11.50	5.00	1.63
					(0.097)	(0.133)
Total forward-looking details	73.05	74.45	76.47	80.33	3.41	6.53
					(0.483)	(0.349)

All p-values are two-tailed.

The information categories are described in the Appendix. The significance of mean pair-wise differences is based on one-sample t-tests and the significance of median pair-wise differences is based on the Wilcoxon sign-rank test.

analysis, we find that the total volume of information for the conference call transactions is eight times as large as the total volume of information for the non-conference call transactions. Moreover, the significantly greater volume of disclosure in conference call transactions extends to all information categories.

Table 9 reports between-transaction comparisons of the relative emphasis on information categories in merger announcement press releases versus conference call transcripts. While roughly two-thirds of the disclosure volume in conference call transactions is devoted to forward-looking information, only one-third of the disclosure volume in control transactions is devoted to forward-looking information. Looking at individual categories, we find that control transactions emphasize information on deal details and the acquirer's business more than the conference call transactions, while conference call transactions emphasize discussion of the strategic rationale, integration details, and financial guidance more than the control transactions. Overall, the emphasis on forward-looking information in the conference call transactions is approximately 30 percent higher than that in the control transactions ($p < 0.001$ for the mean and median pairwise differences).

Association between Differences in Information Content and Differences in Market Reaction

The foregoing content analysis indicates that conference call transactions provide a greater volume of disclosure and emphasize forward-looking details to a greater degree. To determine

TABLE 9

Pairwise-Comparisons of the Relative Emphasis on Information Categories for the 20 Stock-for-Stock Merger Announcements Accompanied by Conferences and a Control Sample of 20 Stock-for-Stock Transactions with a Press Release Only

	(1) Conference Call Sample		(2) Control Sample		(1)−(2) Pairwise Difference	
	Mean	Median	Mean	Median	Mean (p-value)	Median (p-value)
Deal details	13.51	11.98	37.76	32.11	−24.45 (<0.001)	−17.46 (<0.001)
Acquirer business	7.04	5.96	13.85	13.94	−6.81 (<0.001)	−7.06 (<0.001)
Target business	13.51	11.77	14.31	12.03	0.80 (0.835)	0.50 (0.870)
Strategic rationale	40.69	38.48	24.71	26.07	15.98 (0.001)	15.51 (0.001)
Integration details	16.51	13.59	5.65	2.86	10.86 (0.001)	10.83 (0.001)
Financial guidance	8.73	8.51	3.71	2.40	5.02 (<0.001)	4.99 (<0.001)
Total forward-looking details	65.93	69.33	34.07	33.21	31.86 (<0.001)	30.63 (<0.001)

All p-values are two-tailed.

The information categories are described in the Appendix. The significance of mean pair-wise differences is based on one-sample t-tests and the significance of median pair-wise differences is based on the Wilcoxon sign-rank test.

whether these differences in information content are indeed driving the relatively more favorable response to conference call transactions, we use the following two alternative measures of information content:

DIFFWORDCOUNT = the ratio of the total word count provided in conference call transactions to the total word count provided in control transactions; and

DIFFFUTURE = the percentage of total words devoted to forward-looking details provided in conference call transactions minus the percentage of total words devoted to forward-looking details provided in the control transactions.

For each information content measure, we designate pairs in the upper (lower) half of the sample distribution as those where the amounts of additional information provided in the conference call transaction relative to the control transaction are large (small). We then examine pairwise differences in announcement period returns for each subsample. The results in Table 10 show that, for both information content measures, the relatively more favorable investor responses to conference call transactions is limited to those transactions where the information differential between the conference call transaction and the control transactions is large. This result provides some indication that the greater volume of details disclosed, as well as the greater emphasis on forward-looking information, contribute to the more favorable reactions observed for the conference call transactions.

VIII. CONCLUSION

We investigate the factors that influence whether bidders engage in voluntary disclosure at the time they announce a proposed transaction. We find that bidders are more likely to hold conference calls in connection with the announcement of proposed mergers as the proportion of consid-

TABLE 10

Analysis of Pairwise Differences in Market Reactions Based on Pairwise Differences in Information Content

Information Content Measure	Large Differences in Information Content	Small Differences in Information Content
DIFFWORDCOUNT	Mean: 8.81% (p = 0.034) Median: 7.41% (p = 0.049) {n = 10}	Mean: 0.77% (p = 0.476) Median: 0.00% (p = 0.625) {n = 10}
DIFFFUTURE	Mean: 6.94% (p = 0.034) Median: 5.05% (p = 0.027) {n = 10}	Mean: 2.65% (p = 0.382) Median: 0.85% (p = 0.625) {n = 10}

All p-values are two-tailed.

We rank our 20 pairs of conference call-control transactions on the basis of differences in information content (which we measure alternatively as *DIFFWORDCOUNT* and *DIFFFUTURE*). We designate those pairs that are above (below) the median as having large (small) differences in information content.

DIFFWORDCOUNT is the ratio of the total word count provided in conference call transactions to the total word count provided in control transactions. *DIFFFUTURE* is the percentage of total words devoted to forward-looking details provided in conference call transactions minus the percentage of total words devoted to forward-looking details provided in the control transactions.

eration to be paid in the form of the bidder's stock and the relative size of the deal increase. After controlling for endogeneity, we find that the average market reaction to merger announcements that are accompanied by concurrent conference calls is significantly more favorable than the average market reaction to announcements where the acquirers rely exclusively on press releases to introduce the deals. Based on a content analysis of a limited subsample of stock acquirers, we provide evidence that the greater volume of information provided in conference calls relative to press releases, as well as the greater emphasis on forward-looking details in conference calls relative to press releases, contributes to the more favorable returns to merger announcements that are accompanied by conference calls. We find no evidence that the superior announcement returns subsequently reverse or that conference calls are associated with pre-announcement abnormal accruals.

Overall, our study demonstrates that bidders use merger-related supplemental disclosure at the time of the merger announcement as a means to positively influence their stock prices. In addition, we document the usefulness of voluntary disclosure in the context of a significant strategic decision. More importantly, our analysis suggests that, in general, managers use conference calls around merger announcements as a means of conveying genuine private information as opposed to falsely hyping their stock prices.

APPENDIX
TAXONOMY USED TO CLASSIFY PRESS RELEASE AND CONFERENCE CALL DISCLOSURES

Historical Details	Forward-Looking Details
Deal Details	**Strategic Rationale**
Amount and form of consideration	Enumeration of strategic benefits/source of synergies
Signing of definitive agreement	Factors impacting strategy's success
Determination and justification of purchase price	Merger as part of overall strategic vision
Due diligence procedures/negotiations	Actions taken or to be taken to implement strategy
Additional details (collars, breakup fees, earnouts, etc.)	General statements of optimism
Funding of merger	Customer overlap/new customer base
Regulatory and other considerations	Market expansion
Expected time to completion	Product focus/distribution channel
Approval status of target shareholders	Position of combined entity in industry
Financial advisor	Quantification of cost savings
Merger premium	Impact of merger on partners and customers
Approval status of Board of Directors	Plans for future business combinations
Historical relation between acquirer and target	Impact of transaction on other strategic initiatives
Other bidders	**Integration Details**
Related fees and charges	General discussion
Approval status of acquirer shareholders	Technological integration
Acquirer Business	Plans for facilities
Description of business	Personnel integration
Business model	Cultural fit
Description of products	Fate of top management
Description of acquirer assets	Discussion of the team responsible for integration
Customer base	Past successes in integration
Historical financial results	**Financial Guidance**
Historical stock price performance	Guidance about income statement impact
Historical nonfinancial metrics	Accretive/decretive
Target Business	Hypothetical combined results
Description of business	Qualitative guidance

(continued on next page)

Historical Details	Forward-Looking Details
Target Business	**Financial Guidance**
Description of target assets	Effect of merger on nonfinancial measures
Description of products	Update of previously issued guidance
Target's historical financial results	Accounting for combination
Business model	Hypothetical combined balance sheet
Market position	Capital expenditure forecast
Management team/employees	Sarbanes-Oxley/SEC reviews
Customer base	Capital structure of combined entity
Revenue mix	Impact on key ratios/rate of return estimates
Strategic partnerships	Impact on specific segments

REFERENCES

Ahearne, A., W. Griever, and F. Warnock. 2004. Information costs and home bias: an analysis of U.S. holdings of foreign securities. *Journal of International Economics* 62 (2): 313–336.

Asquith, P., R. Bruner, and D. Mullins, Jr. 1983. The gains to bidding firms from mergers. *Journal of Financial Economics* 11 (1–4): 121–139.

Barber, R., and J. Lyon. 1997. Detecting long-run abnormal stock returns: the empirical power and specifications of test statistics. *Journal of Financial Economics* 43 (3): 341–372.

Baginski, S., and J. Hassell. 1997. Determinants of management forecast precision. *The Accounting Review* 72 (2): 302–312.

Bowen, R., A. Davis, and D. Matsumoto. 2002. Do conference calls affect analysts' forecasts? *The Accounting Review* 77 (2): 285–316.

Bradshaw, M., B. Bushee, and G. Miller. 2004. Accounting choice, home bias, and U.S. investment in non-U.S. firms. *Journal of Accounting Research* 42 (5): 795–841.

Brockman, P., and X. Martin. 2009. Voluntary disclosures around acquisitions. Working paper, University of Missouri, and Washington University in St. Louis.

Chang, S. 1998. Takeovers of privately held targets, method of payment, and bidder returns. *The Journal of Finance* 53 (2): 773–784.

Chung, K., R. Jacob, and Y. Tang. 2003. Earnings management by firms announcing earnings after SEC filing. *International Advances in Economic Research* 9 (2): 152–162.

Covrig, V., S. Lau, and L. Ng. 2006. Do domestic and foreign fund managers have similar preferences for stock characteristics? A cross-country analysis. *Journal of International Business Studies* 37 (3): 407–429.

Cready, W., and P. Mynatt. 1991. The information content of annual reports: A price and trading response analysis. *The Accounting Review* 66 (2): 291–312.

Dye, R. 1985. Disclosure of nonproprietary information. *Journal of Accounting Research* 23 (1): 123–145.

Easton, P., and M. Zmijewski. 1993. SEC Form 10 K/10Q reports and annual reports to shareholders: Reporting lags and squared market model prediction errors. *Journal of Accounting Research* 31 (1): 113–129.

Erickson, M., and S. Wang. 1999. Earnings management by acquiring firms in stock for stock mergers. *Journal of Accounting and Economics* 27 (2): 149–176.

Fischer, P., and H. Louis. 2008. Financial reporting and conflicting managerial incentives: The case of management buyouts. *Management Science* 54 (10): 1700–1714.

Foster, T., D. Jenkins, and D. Vickery. 1983. Additional evidence on the incremental information content of the 10-K. *Journal of Business Finance & Accounting* 10 (1): 57–66.

Francis, J., D. Philbrick, and K. Schipper. 1994. Shareholder litigation and corporate disclosures. *Journal of Accounting Research* 32 (2): 137–164.

Francis, J., K. Schipper, and L. Vincent. 2002. Expanded disclosures and the increased usefulness of earnings announcements. *The Accounting Review* 77 (3): 515–546.

Frankel, R., M. Johnson, and D. Skinner. 1999. An empirical examination of conference calls as a voluntary

disclosure medium. *Journal of Accounting Research* 37 (1): 133–150.

Fried, D., and A. Schiff. 1981. CPA switches and associated market reactions. *The Accounting Review* 56 (2): 326–331.

Fuller, K., J. Netter, and M. Stegemoller. 2002. What do returns to acquiring firms tell us? Evidence from firms that make many acquisitions. *The Journal of Finance* 57: 1763–1793.

Gong, G., H. Louis, and A. Sun. 2008a. Earnings management, lawsuits, and stock-for-stock acquirers' market performance. *Journal of Accounting and Economics* 46 (1): 62–77.

———, ———, and ———. 2008b. Earnings management and firm performance following open-market repurchases. *The Journal of Finance* 63 (2): 947–986.

Griffin, P. 2003. Got information? Investor response to Form 10-K and Form 10-Q EDGAR filings. *Review of Accounting Studies* 8 (4): 433–460.

Heckman, J. 1979. Sample selection bias as a specification error. *Econometrica* 47 (1): 153–161.

Hollie, D., J. Livnat, and B. Segal. 2005. Oops, our earnings were indeed preliminary: When preliminary earnings announcements and subsequent filings disagree. *Journal of Portfolio Management* 31 (2): 94–104.

Hutton, A., G. Miller, and D. Skinner. 2003. The role of supplementary statements with management earnings forecasts. *Journal of Accounting Research* 41 (5): 867–890.

Jo, H., and Y. Kim. 2007. Disclosure frequency and earnings management. *Journal of Financial Economics* 84 (2): 561–590.

Johnson, W., and T. Lys. 1990. The market for audit services: Evidence from voluntary auditor changes. *Journal of Accounting and Economics* 12 (1–3): 281–308.

Jones, J. 1991. Earnings management during import relief investigations. *Journal of Accounting Research* 29 (2): 193–228.

Jung, W., and Y. Kwon. 1988. Disclosure when the market is unsure of information endowment of managers. *Journal of Accounting Research* 26 (1): 146–153.

Kang, J., and R. Stulz. 1997. Why is there a home bias? An analysis of foreign portfolio equity ownership in Japan. *Journal of Financial Economics* 46: 3–28.

Kasznik, R., and B. Lev. 1995. To warn or not to warn: Management disclosures in the face of an earnings surprise. *The Accounting Review* 70 (1): 113–134.

Kimbrough, M. 2005. The effect of conference calls on analyst and market underreaction to earnings announcements. *The Accounting Review* 80 (1): 189–219.

Klock, M. 1994. The stock market reaction to a change in certifying auditor. *Journal of Accounting, Auditing & Finance* 9 (2): 339–347.

Kothari, S., A. Leone, and C. Wasley. 2005. Performance matched discretionary accruals. *Journal of Accounting and Economics* 39 (1): 163–197.

———, and J. Warner. 1997. Measuring long-horizon security price performance. *Journal of Financial Economics* 43 (3): 301–339.

Lang, L., R. Stulz, and R. Walkling. 1989. Managerial performance, Tobin's q and the gain from successful tender offers. *Journal of Financial Economics* 24 (1): 137–154.

Lang, M., and R. Lundholm. 2000. Voluntary disclosure and equity offerings: Reducing information asymmetry or hyping the stock? *Contemporary Accounting Research* 17 (4): 623–662.

Larcker, D., and T. Rusticus. 2010. On the use of instrumental variables in accounting research. *Journal of Accounting and Economics* 49 (3): 186–205.

Leone, A., S. Rock, and M. Willenborg. 2007. Disclosure of intended use of IPO proceeds and underpricing in initial public offerings. *Journal of Accounting Research* 45 (1): 111–153.

Li, E., and K. Ramesh. 2009. Market reaction surrounding the filing of periodic SEC reports. *The Accounting Review* 84 (4): 1171–1208.

Lipin, S., and M. Sirower. 2003. Investor communications: A message of confidence in the deal. *Mergers and Acquisitions* 38 (10): 21–26.

Loderer, C., and K. Martin. 1997. Executive stock ownership and performance: Tracking faint traces. *Journal of Financial Economics* 45: 223–255.

Louis, H. 2004. Earnings management and the market performance of acquiring firms. *Journal of Financial Economics* 74 (1): 121–148.

———, D. Robinson, and A. Sbaraglia. 2008. An integrated analysis of the association between accrual disclosure and the abnormal accrual anomaly. *Review of Accounting Studies* 13 (1): 23–54.

———, and H. White. 2007. Do managers intentionally use repurchase tender offers to signal private information? Evidence from firm financial reporting behavior. *Journal of Financial Economics* 85 (1): 205–233.

Malmendier, U., and G. Tate. 2008. Who makes acquisitions: CEO overconfidence and the market's reaction. *Journal of Financial Economics* 89 (1): 20–43.

Maquieira, C., W. Megginson, and L. Nail. 1998. Wealth creation versus wealth redistributions in pure stock-for-stock mergers. *Journal of Financial Economics* 48 (1): 3–33.

Matsumoto, D., M. Pronk, and E. Roelofson. 2006. Managerial disclosure vs. analyst inquiry: An empirical investigation of the presentation and discussion portions of earnings-related conference calls. Working paper, University of Washington, and Erasmus University Rotterdam.

Mayew, W., and M. Venkatachalam. 2009. The power of voice: Managerial affective states and future firm performance. Working paper, Duke University.

Moeller, S., F. Schlingemann, and R. Stulz. 2005. Wealth destruction on a massive scale? A study of acquiring-firm returns in the recent merger wave. *The Journal of Finance* 60 (2): 757–782.

———, ———, and ———. 2007. How do diversity of opinion and information asymmetry affect acquirer returns? *Review of Financial Studies* 20 (6): 2047–2078.

Moore, W., E. Rimland, H. Ritch, and J. Rouner. 1998. On the acquisition fast-track to growth and value. *Mergers and Acquisitions* 32 (4): 10–17.

Morck, R., A. Shleifer, and R. Vishny. 1990. Do managerial objectives drive bad acquisitions? *The Journal of Finance* 45: 31–48.

Pastena, V. 1979. Some evidence on the SEC's system of continuous disclosure. *The Accounting Review* 54 (4): 776–783.

Schwartz, K., and B. Soo. 1995. An analysis of Form 8-K disclosures of auditor changes by firms approaching bankruptcy. *Auditing: A Journal of Practice & Theory* 14 (1): 125–136.

Securities and Exchange Commission (SEC). 2000. Final rule: Selective disclosure and insider trading. Available at http://www.sec.gov/rules/final/33-7881.htm.

Servaes, H. 1991. Tobin's Q and the gains from takeovers. *The Journal of Finance* 46: 409–419.

Shleifer, A., and R. Vishny. 2003. Stock market driven acquisitions. *Journal of Financial Economics* 70 (3): 295–311.

Skinner, D. 1997. Earnings disclosure and stockholder lawsuits. *Journal of Accounting and Economics* 23 (3): 249–282.

Stice, E. 1991. The market reaction to 10-K and 10-Q filings and to subsequent The Wall Street Journal earnings announcements. *The Accounting Review* 66 (1): 42–55.

Suh, J. 2001. Home bias among institutional investors: A study of *The Economist* quarterly portfolio poll. Working paper, SK Economic Research Institute.

Tasker, S. 1998. Bridging the information gap: Quarterly conference calls as a medium for disclosure. *Review of Accounting Studies* 3 (1–2): 137–167.

Travlos, N. 1987. Corporate takeover bids, method of payment, and bidding firms' stock returns. *The Journal of Finance* 42 (4): 943–963.

Tucker, J. 2007. Is openness penalized? Stock returns around earnings warnings. *The Accounting Review* 82 (4): 1055–1087.

Wooldridge, J. 2002. *Econometric Analysis of Cross-Section and Panel Data*. Cambridge, MA: The MIT Press.

———. 2008. Estimation of the average treatment effect in the correlated random coefficient model. In *Advances in Econometrics* Vol. 21, *Modeling and Evaluating Treatment Effects In Econometrics*, edited by D. Millimet, J. Smith, and E.. Vytlacil, 93–117. Amsterdam, The Netherlands: Elsevier.

THE ACCOUNTING REVIEW
Vol. 86, No. 2
2011
pp. 669–701

American Accounting Association
DOI: 10.2308/accr.00000023

The Role of Newswires in Screening and Disseminating Value-Relevant Information in Periodic SEC Reports

Edward Xuejun Li
University of Rochester

K. Ramesh
Rice University

Min Shen
George Mason University

ABSTRACT: We examine the role of newswires in identifying and conveying market-moving information in periodic SEC reports to capital market participants. Using data on Dow Jones Newswires, we find that newswires are more likely to send alerts on firms that do not release preliminary earnings, have credit ratings, are included in major market indices, have litigation exposure, or report losses. Reflective of the market's focus on certain key events, firms with a nonstandard audit opinion, in the process of delisting, reporting unusual accounting items, or raising equity capital also receive alerts. Moreover, not only do we find significant price and volume reactions to the alerts at the daily level, but also we document immediate intra-day market activity triggered by the alerts, whereas we detect no similar reaction for SEC filings that trigger the alerts. Additional analysis suggests that the intra-day reaction is not driven by noise trading.

Keywords: *information intermediaries; newswires; periodic SEC reports; information dissemination; market microstructure.*

The authors benefited from discussions with Rick Stine (senior editor, Editorial Operations for the Americas, Dow Jones Newswires), Matt Kalan of Progress Software, and Herbert Lichtman of Arena Consulting LLC. Comments and suggestions of Shannon Anderson, Jim Brickley, Steven Kachelmeier (senior editor), Christian Leuz, Andy Leone, Karen Nelson, Sundaresh Ramnath, James Sinclair, Mark Trombley (editor), Peter Wysocki, Steve Zeff, Jerry Zimmerman, two anonymous reviewers, and other workshop participants at Rice University, Syracuse University, University of Miami, University of Rochester, and the University of Utah Winter Accounting Conference (especially the discussant, Eugene Soltes) are gratefully acknowledged. We thank Novia Chen for her research assistance. We are responsible for all remaining errors. Professor Lynn M. LoPucki generously provided the bankruptcy data used in this study. Min Shen gratefully acknowledges the summer research grant from the School of Management and the Provost Office of George Mason University.

Part of the research was completed when K. Ramesh was an Academic Fellow at the Securities and Exchange Commission, which, as a matter of policy, disclaims responsibility for any private publication or statements by any of its employees or contractors. The views expressed herein are those of the authors and do not necessarily reflect those of the Commission or its staff.

Editor's note: Accepted by Mark Trombley.

Submitted: November 2009
Accepted: August 2010
Published Online: March 2011

Data Availability: *Various data sources are identified in the text.*

I. INTRODUCTION

This study examines the role of newswires as an information intermediary in capital markets. We focus on Dow Jones Corporate Filing Alert (hereafter, DJCFA), a service offered by Dow Jones Newswires that screens for and disseminates important market-moving information in SEC filings to professional investors and other market participants in a timely manner.[1] What firm and informational characteristics in a periodic report would motivate a newswire service to issue an alert? Given that the SEC filings are already publicly available on EDGAR, is the information disseminated by the newswire service incrementally value-relevant to investors? We address these two questions in this study.

Despite extensive research on the role of financial analysts in the capital market, there is only limited evidence on the economic function of *pure* information intermediaries such as newswires.[2] We choose a context in which information intermediaries must be highly skilled to identify market-moving information. As earnings press releases become increasingly preemptive (Francis et al. 2002; D'Souza et al. 2010a), swift identification of any value-relevant information in periodic reports requires astute information screening by intermediaries. In addition, despite the sophistication of U.S. capital markets, academic research suggests that the confluence of demand- and supply-side effects restricts the market's processing ability when it is inundated with information or when facing complex information (Hirshleifer et al. 2008; You and Zhang 2009), which is commonly the case in periodic reports (Griffin 2003; Li and Ramesh 2009a).

We posit that even large investors generally do not satisfy the cost-benefit calculus for combing through periodic reports of every stock they own or in their investment possibility set to identify market-moving information. The broad market demand for corporate accounting information, coupled with the prevalence of high fixed costs and close-to-zero marginal costs in the market for information goods (Romer 1990), provides the impetus for information intermediaries such as DJCFA to enter the marketplace. To quote Dow Jones:

> From initial public offerings to Chapter 7 liquidations, Dow Jones Corporate Filings Alert offers readers invaluable information and analysis beyond news available in press releases. As the volume of company filings to the SEC keeps growing, we can save our readers time and give them the edge they need in an information-overloaded environment. ("Dow Jones Newswires Re-Launches Corporate Filings Alert; Division Will Also Open First Bureau in Wilmington, Delaware," April 2, 2001, Business Wire)

Our analysis is based on a comprehensive sample of 26,615 alerts pertaining to 20,797 periodic SEC reports issued by Dow Jones over the post-EDGAR period 1997–2004, representing 9 percent of all periodic reports in our sample. While roughly 82 percent of the filings with alerts

---

[1] The alerts we examine are mostly recitation of specific facts already disclosed by companies in their periodic SEC reports and, therefore, are unlikely to reflect the *quid pro quo* relation found between reporters and their corporate sources (Dyck and Zingales 2003). We understand that Dow Jones News Services has strict conflict-of-interest policies to ensure independence in coverage of companies. Also, given the factual nature of the Dow Jones alerts, they do not suffer from the spin effect presented in the typical news stories (Dyck and Zingales 2003; In et al. 2007).

[2] We view a pure information intermediary as "an independent, profit-maximizing information processing system performing its activities (information acquisition, processing, and distribution) on behalf of other agents' information needs" (Rose 1999, 79). Academic research on newswires as an information intermediary include Dawkins and Bamber (1998) who find that the adverse market reaction to corporate bankruptcy filings is mostly due to information in bankruptcy filings disseminated through Broad Tape, and Dopuch et al. (1986) who find that the stock market reacts adversely only to audit qualifications discussed by media. Also see Dyck and Zingales (2003), Bhattacharya et al. (2009), Fang and Peress (2009), Solomon (2009), Soltes (2009), and Bushee et al. (2010) on the broader role of business press in shaping the capital market information environment. See D'Souza et al. (2010b) on the role of capital market data aggregators.

receive only one alert, another 12 percent receive two, with the remainder receiving three or more. The number of news categories in an alert ranges from one to seven, with a mean value of two.[3] While Dow Jones takes 2.3 weekdays on average to release an alert after the corresponding filing, 68 percent of the initial alerts are released within 24 hours.

Our first analysis examines the demand-and-supply considerations in the information market that guide the alert decision. Specifically, we posit that the market demand for information arises from investor awareness and the information environment. Consistent with the investor awareness argument (Wurgler and Zhuravskaya 2002; Chen et al. 2004; Chen et al. 2006; Docking and Dowen 2006; Elliott et al. 2006; Mase 2006; D'Souza et al. 2010b), firms included in S&P major market indices are more likely to receive a Dow Jones alert. Reflective of equity and credit market demands, analyst following and coverage by credit-rating agencies are positively associated with the incidence of receiving alerts. Investors in loss firms are alerted more often, given that traditional profitability measures alone are insufficient for valuation purposes (Hayn 1995; Collins et al. 1997; Chen et al. 2002; Ajinkya et al. 2005). Firms skipping preliminary earnings announcements also experience a higher incidence of alerts, given that their periodic reports are more apt to include information hitherto not disclosed to the marketplace. Consistent with the value of identifying market-moving information, we find more alerts on firms from industries with higher litigation risk.

Among the key firm-specific events, firms raising equity capital, approaching delisting, experiencing large price reactions at preliminary earnings announcements, reporting extraordinary or special items, or receiving a nonstandard audit opinion are more likely to prompt an alert by Dow Jones. In terms of economic significance, nonstandard audit reports increase the probability of an alert by 5.1 percentage points, followed by firms without a preliminary earnings announcement (3.7 percentage points), periodic reports of firms with a credit rating (3.0 percentage points), firms subject to delisting (1.9 percentage points), firms listed in major market indices (1.2 percentage points), and firms facing a higher litigation risk (1.1 percentage points). In terms of supply-side effects, we find that Dow Jones' propensity to issue an alert drops when the market is inundated with a large number of periodic reports. We also provide evidence that Dow Jones switches resources away from periodic reports on days when there is increased incidence of earnings releases in the marketplace. Conditional on issuing an alert, we find that the time lag between issuing the initial alert and the filing of the periodic report is roughly 40 (22) percent shorter in firm-quarters without a preliminary earnings announcement (for firms in industries with higher litigation risk).

Our second set of analyses focuses on the information content of filing alerts. After controlling for earnings announcements and filing of periodic reports, we find significant price and volume reactions on the days when Dow Jones issues an alert. However, the reactions are limited to the initial alert regardless of whether the initial alert is followed by another alert. Using the *Alert Day* observations, we also find that, while legal- and bankruptcy-related alerts consistently generate higher price and volume reactions compared to the other news categories, alerts with missing subject codes generate significantly lower market reactions, consistent with constraints on immediate trading either based on algorithms or that require human intervention.

One obvious concern with using daily data is the possible endogeneity that market participants are more focused on filings that trigger an alert for reasons other than the information contained in the alerts themselves. We address this endogeneity by comparing the price and volume reactions over a narrow intra-day period surrounding the release of filing alerts with those

[3] Dow Jones categorizes the alerts based on more than 200 subject codes, which we collapse into ten broad news categories. See Section III for details.

of a pseudo-event matched on the weekday and release time of the alert. Compared to the 60-minute pre-event period, the absolute stock returns (trading volume) for each of the first five minutes following the alert are between 27 and 52 percent (18 and 69 percent) higher, with no similar reactions following pseudo-events. In addition, our formal statistical tests show that the price and volume reactions to alerts during the first 16 minutes (including the event minute) are significantly higher than those during the comparable pre- and post-event windows as well as those during the corresponding pseudo-event window. However, when we base our analysis on the actual filing time of periodic reports, we find no significant price or volume reactions. Our results for the alerts and filings hold when we split the sample observations into 10-K filings versus 10-Qs. More importantly, despite the fact that there are abnormally low price and volume reactions generally surrounding the release of 10-Q reports (Li and Ramesh 2009a), 10-Q-based alerts generate immediate price and volume reactions comparable to those generated by10-K-based alerts. In addition, we find significant volume reactions from both large and small trades to alerts, with small trades exhibiting a more pronounced volume activity. Finally, we conduct a battery of tests to show that the intra-day price/volume reactions to alerts are not merely due to noise trading or market over-reaction.

Overall, our study makes several contributions to the extant literature. First, capital allocation and social welfare are clearly dependent on the efficiency of financial markets (Jennings and Barry 1983). Our study provides large-sample evidence regarding the key role played by newswires in reducing information overload faced by market participants and improving price efficiency. Second, while the traditional asset-pricing literature assumes instantaneous information diffusion and complete market reaction to publicly available information (Merton 1987), the behavioral finance literature allows for bounded rationality and slow diffusion of information (e.g., Hong and Stein 1999). However, there is no substantial body of literature that examines the cost-benefit trade-offs market participants face in shaping the capital market information infrastructure. Our study helps to fill this gap by showing that newswire services such as DJCFA act as delegated information intermediaries in the capital market. Finally, our study is an early attempt in using market microstructure data to study the role of information intermediaries. This approach has broader application for parsing the pricing and trading implications in numerous contexts with confounding events (Brown et al. 1992; Ecker et al. 2006; Li and Ramesh 2009a).

Section II discusses the role of newswires in shaping the capital market information infrastructure and describes the Dow Jones alert service. Section III describes our sample screening procedures and presents descriptive evidence on the filing alerts. The analyses of factors that influence Dow Jones to issue an alert and the market reaction to the alerts are detailed in Sections IV and V, respectively. The final section provides a conclusion.

II. NEWSWIRES AND THE MARKET FOR INFORMATION IN PERIODIC REPORTS

In this section, we first postulate the role of newswires generally in the market for information in periodic SEC reports, followed by a discussion of the Dow Jones Corporate Filing Alert service, the specific newswire service that we study.

Role of Newswires in the Market for Information in Periodic SEC Reports

Consistent with earnings releases increasingly preempting periodic SEC reports (Francis et al. 2002; D'Souza et al. 2010a), Li and Ramesh (2009a) report no significant price or volume reactions surrounding the filing of quarterly periodic reports and find evidence of market reaction to

10-K reports only under limited circumstances.[4] When periodic reports contain key value-relevant information, in some circumstances managers themselves have incentives to highlight it to mitigate litigation risk (Li and Ramesh 2009a; Li and Ramesh 2009b). However, when such disclosure incentives are absent, the speed with which the market assimilates new information from periodic SEC reports must be determined by the demand and supply forces at work in the market for corporate accounting information. Recent research suggests that when corporate information releases cluster in calendar time (Hirshleifer et al. 2008) or contain extensive textual information as in periodic SEC reports (You and Zhang 2009), the stock market does not process the information in a timely fashion.

However, while most investors acting alone would find it formidable to perform detailed quantitative and textual analysis on periodic reports of all stocks in their investment set, the nature of market for information goods suggests that information intermediaries such as newswires would enter the marketplace acting as information-gathering agents and serve a multitude of principals. Romer (1990) suggests that, although the market for information goods is characterized by large fixed costs, it faces trivial reproduction and dissemination costs, thereby virtually eliminating marginal cost. Given the broad market demand for corporate accounting information, sophisticated information intermediaries can obtain economies of scale by spreading the large fixed cost of information retrieval and processing among many customers (Veldkamp 2006). We argue that these market characteristics provide the impetus for newswire services with extensive expertise in the financial information market to act as delegated information intermediaries to large groups of market participants for collecting key information from periodic reports.

Description of Dow Jones Corporate Filing Alert Service

Newswire services facilitate both the corporate supply of accounting information and the subsequent retrieval by various market participants. While Business Wire and PR Newswire act as agents of public corporations in the information-dissemination process, the major wire services such as Dow Jones, Reuters, and Bloomberg cater to user needs through sophisticated real-time delivery of business news and news alerts. We choose to examine the DJCFA of Dow Jones, Inc. for three reasons. First, Dow Jones is a leading provider of business content and information services to the capital markets through newswires and other channels. Second, DJCFA specifically targets SEC filings, the focus of this study. Third, the alerts issued by this service are publicly available through Factiva. Below we provide a brief description of the history and the information services provided by DJCFA.[5]

In 1989, Dow Jones acquired Federal Filings Inc., a private company that acted as an information intermediary by culling through the voluminous SEC paper filings to extract and disseminate key information to various capital market participants including Wall Street. Dow Jones initially provided the alerts under Federal Filings Newswires, which was relaunched as DJCFA in 2001. The purpose of the alert service is to screen for key information disclosed in SEC filings that were not preceded by a press release, identify various value-relevant information buried in SEC filings, and to offer market-moving alerts. Given that selective disclosure is prohibited under Reg FD, DJCFA monitors SEC filings to quickly draw professional investors' attention to key corporate disclosures in the post-Reg FD period.

---

[4] Beyer et al. (2009) find that market reactions to voluntary disclosures are in general greater than that for mandatory disclosures, which suggests that periodic reports may play more of a confirmatory role, adding credibility to prior voluntary disclosures such as preliminary earnings announcements or management forecasts.

[5] This sub-section is based on our discussions with Rick Stine of Dow Jones, various news releases relating to DJCFA obtained through Factiva, online search, and other publicly available news sources. We limit our discussions to SEC filings although DJCFA also reviewed bankruptcy court filings and information sources.

During the period we study, DJCFA was staffed by editors and reporters with industry-specific expertise to search for market-moving items of interest to professional investors and Wall Street. Specifically, the DJCFA staff determines data-collection strategies and dissemination speed based on perceived market following and news value.[6]

The DJCFA service targets traders on the floor of major exchanges, at the trading desks of various brokerage houses, mutual funds, and hedge funds, although individual investors can access a large portion of real-time news from Dow Jones Newswires through *The Wall Street Journal Online* (http://www.wsj.com). Along with DJCFA, Dow Jones news is distributed to over 100 vendors. Market participants who need real-time information obtain Dow Jones news through distribution platforms hosted by Thomson Reuters, Bloomberg, FactSet, etc. Those interested in archival research rely on services such as Factiva.[7]

III. SAMPLE AND DESCRIPTIVE EVIDENCE ON DOW JONES CORPORATE FILING ALERTS

In this section, we first describe our sample construction, and then provide descriptive statistics on Dow Jones Corporate Filing alerts.

Sample Screening Procedure and Distribution

Our sample screening procedure is detailed in Table 1, Panel A. We ran a keyword search in Factiva Academic and obtained 36,984 entries for the period of 1997–2004.[8] Of these, 3,220 are advertisements for the DJCFA service, 1,996 are news articles related to a group of firms rather than an individual firm's periodic report, and 438 are rumors. Given that we focus on the alerts that follow 10-K/10-Q filings, we exclude alerts pertaining to 8-Ks, notices of non-timely filing, amended filings, and periodic filings of foreign companies. We further drop alerts for small business companies because they face different mandatory disclosure standards. Finally, we delete a small set of observations with alerts unrelated to SEC filings (230), alerts related to SEC filings other than 10-Ks/10-Qs (67), or alerts without filing time stamps (54). The final sample consists of 26,615 alerts. Panel B indicates that these alerts are related to 20,797 periodic SEC filings (hereafter, alerted filings). While more than 80 percent of these filings generate only a single alert, a few filings generate as many as five or more alerts.

Untabulated results show that, while the proportion of 10-K/10-Q filings with an alert monotonically declined from 1997 to 2000, it rebounded beginning 2001, reflecting Dow Jones' increased emphasis on the alert service after Reg FD.[9] When examined by industry, the alert sample is representative of the overall Compustat population, except that there are disproportionately more (less) alerted filings in the retail (financial services) industry.

---

[6] While the Dow Jones alert service that we study is largely a recitation of information in periodic SEC reports, the service does require its reporters' judgment to choose what specific information to include in the alert. However, unlike capital market muckrakers (Foster 1979), business press (Miller 2006), or Dow Jones' other alert/newswire services, DJCFA includes neither critical analyses nor reporters' assessment of the market implications of the information in the alert.

[7] Beginning July 2, 2009, Dow Jones has restructured the DJCFA service for various financial reasons given the current market conditions. The restructured service will utilize limited DJCFA reporters along with the entire corporate reporting staff at the New York desk to look for market-moving alerts to be sent either as part of DJCFA or broadly through Dow Jones Newswires.

[8] Based on discussions with Factiva, we used the free text string "10-K" or "10-Q" or "10K" or "10Q" or "10KSB" or "10QSB" or "10-KSB" or "10-QSB" with "Dow Jones Corporate Filing Alerts" as the source. We searched the period beginning in 1997, the first full calendar year after EDGAR became effective, and ending in 2004, the last year for which we are able to consistently identify alerts related to periodic reports through *Factiva Academic*.

[9] See "Dow Jones Newswires Re-Launches Corporate Filings Alert; Division Will Also Open First Bureau in Wilmington, Delaware," April 2, 2001, Business Wire.

TABLE 1

Sample of Dow Jones Corporate Filing Alerts

Panel A: Sample Screening of Dow Jones Corporate Filing Alerts

Dow Jones Corporate Periodic Filings Alerts, 1997–2004	36,984
Less: advertisements	(3,220)
news related to a group of firms[a]	(1,996)
rumors[b]	(438)
Initial Sample of Alerts	31,330
Less: alerts after 8-K filings[c]	(514)
alerts after filing of notice of delay under Rule 12b-25[d]	(1,881)
alerts after amended filings[e]	(233)
alerts for foreign firms who file only 20-F/6-Ks with the SEC	(613)
alerts from SB filings[f]	(1,123)
alerts related to non-filing news[g]	(230)
alerts related to non-periodic filings (S-4, 13D, DEF 14A, etc.)	(67)
alerts without filing time[h]	(54)
Final Sample of Alerts	26,615

Panel B: Frequency of Corporate Filing Alerts of Each Periodic SEC Filing

	Frequency	%	Total
1	17,051	81.99	17,051
2	2,529	12.16	5,058
3	750	3.61	2,250
4	269	1.29	1,076
5	126	0.61	630
Above 5	72	0.35	550
	20,797	100.00	26,615

[a] We identify this type of news by searching the news title for keywords such as "corporate filing alerts: the morning's top news," "federal filings business news," "CFA early summary," "high yield information," "daily market wrap-up," etc.

[b] Keywords to identify this type of news are "notice of rumor" or "daily rumor."

[c] Alerts are included in this group if (1) the source of a corporate filing alert indicates 8-K filings; or (2) the firm filed an "8-K" or "8-K/A" on or one day before the alert day, and no 10-Ks/10-Qs were filed within 20 weekdays prior to the alert day.

[d] Alerts are included in this group if (1) the source of corporate filing alert is "NT 10-K" or "NT 10-Q"; (2) the news title contains keywords such as "delayed," "not timely," "extension to file," "files NT," etc.; or (3) the firm filed a form NT 10-K or NT 10-Q on or a day before the alert day, and no 10-Ks/10-Qs were filed within 20 weekdays prior to the alert day.

[e] Alerts are included in this group if (1) the source of corporate filing alert indicates amended filings; (2) the news title contains keywords such as "files amended"; or (3) the firm filed a "10-K/A," "10-K405/A," or "10-Q/A" on or one day before the alert day, and no 10-Ks/10-Qs were filed within 20 weekdays prior to the date of alert.

[f] Alerts are included in this group if (1) the source of corporate filing alert is "10KSB" or "10QSB"; or (2) the firm filed a form 10KSB or 10QSB form on or a day before the alert day.

[g] These alerts are related to earnings announcements, bankruptcies, Moody's downgrading, lawsuits, etc.

[h] This group consists of 26 alerts with corresponding 10-K/10-Q filings available from EDGAR but not in the SEC filing time database and 28 alerts having no corresponding filings from either the EDGAR or the SEC filing time database.

News Content of Corporate Filing Alerts

To analyze news content, we group the more than 200 subject codes provided by Dow Jones Newswires into ten broad news categories (see note in Table 2). For the 26,013 alerts with

TABLE 2
Descriptive Statistics on Alerts by News Categories

Panel A: Frequency Distribution of the Number of News Categories in Alerts

Number of News Categories	Frequency	%	Total
1	10,875	41.81	10,875
2	8,880	34.14	17,760
3	4,313	16.58	12,939
4	1,517	5.83	6,068
5	355	1.36	1,775
6	60	0.23	360
7	13	0.05	91
Total	26,013	100.00	49,868

Panel B: Frequency Distribution of News Categories

Performance (*PER*)	23,785	47.7%
Credit Related (*CRE*)	8,422	16.9%
Equity Related (*EQU*)	4,711	9.4%
Business Structure Change (*BSC*)	2,993	6.0%
Corporate Legal Issue (*LGL*)	2,776	5.6%
Employment Related (*EMP*)	2,635	5.3%
Corporate Business (*BUS*)	920	1.8%
Forecast (*FOR*)	753	1.5%
Bankruptcy (*BKC*)	528	1.1%
Tax Related (*TAX*)	296	0.6%
Other (*OTH*)	2,049	4.1%
Total	49,868	100.0%

We group the more than 200 news subject codes provided by Dow Jones into ten broad news categories, the details of which are available from the authors. Of the 26,615 filing alerts in Table 1, we exclude 326 alerts with missing subject codes and 276 alerts whose subject codes are too broad to be meaningful (e.g., "Corporate/Industrial News," "Commodity/Financial Market News"). Performance news (*PER*) includes information on sales, earnings, and dividends. Credit (*CRE*) (equity news (*EQU*)) provides information on debt (equity) issuances/changes. Business structure changes (*BSC*) capture major corporate events such as mergers and acquisitions, joint ventures, divestures and spin-offs, while corporate business news (*BUS*) provides details on government and nongovernment contracts, new products/services, intellectual property, licensing agreements, research and development, and outsourcing. Issues such as management turnover, executive compensation, layoffs, or insider stock transactions are classified as employment-related news (*EMP*). In addition, we have news groups corresponding to management forecasts (*FOR*), bankruptcy (*BKC*), legal issues (*LGL*), and tax matters (*TAX*). Subject codes that cannot be easily classified into any of the above news categories are included in "Other" (*OTH*).

non-missing subject codes, we identify 49,868 news categories. Table 2, Panel A shows that the number of news categories assigned to an alert ranges from one to seven, and the majority of alerts have more than one news category. Panel B shows that around 50 percent of categories are performance related, which is not surprising given that the primary objective of Dow Jones Corporate Filing Alerts is to highlight market-moving information. The second and third largest groups are credit- and equity-related news, followed by alerts relating to business structure changes, legal issues, and employment matters. Forecasts, bankruptcy, and tax matters are the smallest groups, accounting in total for only three percent of all news categories.

Untabulated analysis indicates that 10-K alerts are more about credit and equity arrangements, employment-related issues, and bankruptcy than 10-Q alerts, possibly due to business events

related to these news categories occurring more frequently around the finalization of the annual report. In contrast, the higher incidence of performance-related news alerts based on 10-Q reports is consistent with Dow Jones grabbing the market's attention when it is likely to be less attentive (Barber and Odean 2008; Li and Ramesh 2009a). Another piece of untabulated descriptive evidence is that, while filings made within trading hours contain significantly more performance-related news, those made after trading hours contain more news in all other categories with several of the categories showing statistically significant differences. One conjecture is that firms try to avoid market attention to non-performance-related news by filing periodic reports after trading hours.

Timing of Dow Jones Corporate Filing Alerts

We obtain corporate filing alert release date/time from *Factiva Academic* and 10-K/10-Q filing date and time stamp from an EDGAR database. The alert release time stamp represents the time when an alert becomes available to the Dow Jones Corporate Filing Alert subscribers, and the 10-K/10-Q filing time stamp is when the periodic filing first appears in EDGAR.[10] Figure 1 plots the frequency of SEC filings and the corresponding initial alerts in blocks of 30-minute intervals. While the number of filings steadily accumulates over most of the SEC working hours of 8:30 a.m. to 5:30 p.m.,[11] the distribution of alerts is relatively smoother and saddle-shaped, with two small peaks at 11:30 a.m. and 3:30 p.m. Untabulated analysis indicates that the first hump for alerts in Figure 1 is caused by Dow Jones' heightened collection efforts to clear the backlog of periodic SEC reports filed during the previous business day.

We next examine the timeliness of corporate filing alerts by measuring the release lag as the number of weekdays between the periodic SEC report filing time and the release time of the first alert (*RL_WEEKDAY*).[12] Because 68 percent of the initial alerts are released within 24 hours after the SEC report is filed, we also measure the release lag in minutes (*RL_MINUTE*) for such alerts. Table 3, Panel A indicates that, on average, it takes Dow Jones Newswires 2.3 weekdays to release an alert after the corresponding SEC report is filed, with one weekday each for the median and the inter-quartile range. We also present statistics on the release lag for 10-K alerts and 10-Q alerts separately and find that the average *RL_WEEKDAY* is almost the same for the two groups. However, for alerts released within one weekday, the median collection time (*RL_MINUTE*) is 43 percent less for alerts based on 10-Q filings than 10-Ks (113 versus 197 minutes), consistent with increased complexity requiring additional time to review 10-K reports for market-moving information.

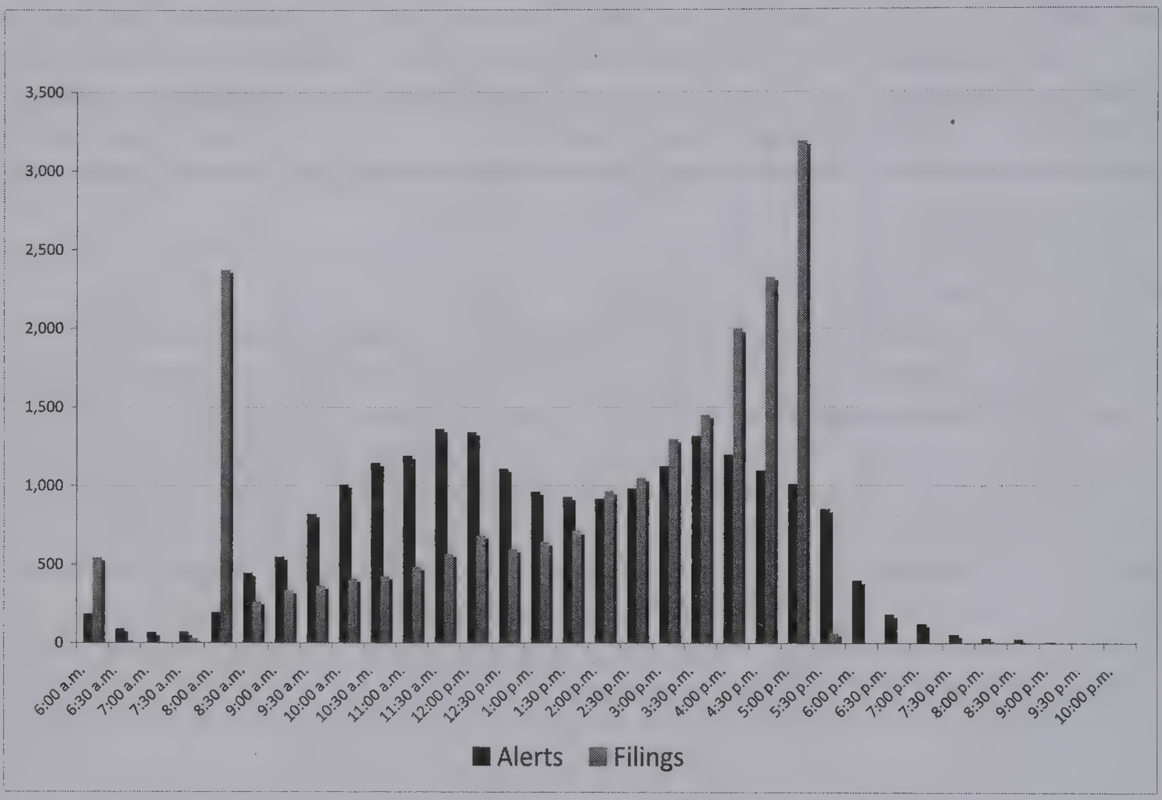
---

[10] Based on our discussions with SEC's Office of Information Technology, we understand that the periodic report time stamp is a good indication of when the filing could be accessed by users, but at least in the earlier years not everyone had instantaneous access to EDGAR (Balsam et al. 2002). When EDGAR was first rolled out, the Level 1 subscribers to EDGAR and their clients (which included many Wall Street firms and investment houses) had immediate access to the filings. The filings were not available at the public website for another 24 hours, which was changed in the year 2002 ("SEC Announces Free, Real-Time Public Access to EDGAR Database," http://www.sec.gov/news/press/2002-75.htm). Untabulated results show that (1) after year 2002 more alerts were issued within the first 15 minutes of filing; and (2) after eliminating these potentially confounding observations from our sample, alerts do (filings do not) trigger significant immediate market activity in periods both before and after the change in the public availability of periodic reports.

[11] Based on our discussions with SEC's Office of Information Technology, periodic reports filed before 5:31 p.m. Eastern time will receive a filing date and time identical to its EDGAR receipt date and time. Filings with a receipt time on or after 5:31 p.m. Eastern time will be assigned the next business day after the receipt date as the filing date and 6:00 a.m. as the filing time. Prior to 2003, the practice appears to have been the use of an 8:00 a.m. time stamp. The two spikes at 6:00 a.m. and 8:00 a.m. are consistent with the SEC time-stamping practice.

[12] We calculate *RL_WEEKDAY* as the number of hours elapsed from the time when a filing appears on EDGAR to the time when the first alert is released by DJCFA and divide it by 24. We subtract 2 if this timeframe contains a weekend. Similarly, the weekend is excluded in computing *RL_MINUTE*.

FIGURE 1
Calendar Time Frequency of SEC Filings and the Corresponding Initial Dow Jones Alerts



This figure presents the time distribution of 20,797 periodic SEC filings and the corresponding initial alerts between 1997 and 2004. The x-axis depicts the time in 30-minute intervals, and the y-axis is the number of periodic SEC filings and the corresponding initial alerts. Unreported analysis reveals similar patterns between 10-K and 10-Q filings/alerts.

Finally, we use the number of news categories in an alert to proxy for the news value to DJCFA subscribers and provide descriptive statistics on the release time of the first alert in Table 3, Panel B. We find that the mean of *RL_WEEKDAY* (the proportion of alerts released within 24 hours after filing) decreases (increases) almost monotonically with the number of news categories, consistent with Dow Jones staff placing higher priority on periodic reports with possibly multiple market-moving information.[13]

IV. DETERMINANTS OF NEWSWIRE ALERTS ON PERIODIC SEC REPORTS

In this section, we first develop predictions regarding the circumstances in which a newswire service is likely to issue an alert based on information in periodic SEC reports. Past research has not identified contexts in which information intermediaries more saliently disseminate to market participants information already disclosed by companies (Dopuch et al. 1986). Our predictions are

[13] Our inferences from Table 3 are unaltered when we exclude observations relating to filings with a time stamp in the interval [6:00 a.m., 8:00 a.m.] Eastern time.

TABLE 3

Time Lag between Periodic SEC Filings and the Initial Corporate Filing Alerts

Panel A: Descriptive Statistics on Time Lag of 10-Ks versus 10-Qs

	RL_WEEKDAY			*RL_MINUTE*		
	Total	**10-K**	**10-Q**	**Total**	**10-K**	**10-Q**
n	20,797	6,929	13,868	14,069	4,632	9,437
Mean	2.325	2.325	2.324	453.187	513.396	423.634
Std.	3.269	3.350	3.227	516.042	520.629	511.215
Min	1	1	1	0	0	0
P5	1	1	1	8	12	7
Q1	1	1	1	40	61	34
Median	1	1	1	137	197	113
Q3	2	2	2	1,064	1,091	1,045
P95	8	7	8	1,337	1,344	1,331
Max	42	42	33	1,440	1,439	1,440

Panel B: Descriptive Statistics on Collection Speed by the Number of News Categories

Number of News Categories	**Obs.**	**Mean of *RL_WEEKDAY***	**Standard Deviation of *RL_WEEKDAY***	**Proportion of One-Weekday Release**
Missing	245	3.294	3.678	0.501
1	8,163	2.338	3.572	0.718
2	6,991	2.463	3.270	0.631
3	3,511	2.245	2.979	0.662
4	1,256	1.823	2.157	0.722
5	302	1.798	2.316	0.722
6	54	1.222	0.945	0.926
7	13	1.154	0.376	0.846

Panel A presents descriptive statistics on the release lag between the periodic SEC report filing time and the release time of the initial alert. *RL_WEEKDAY* is the release lag measured in weekdays, which is calculated as the number of hours elapsed from the time when a filing appears on EDGAR to the time when the first alert is released by DJCFA, divided by 24. Given that 68 percent of alerts are released within 24 hours after the corresponding periodic SEC reports are filed, we also calculate release lag in minutes, denoted *RL_MINUTE*, for these alerts. For both *RL_WEEKDAY* and *RL_MINUTE* we exclude weekends in the calculation. For 74 observations, their filing time stamp is post-alert, which is likely caused by data error. We replace the negative time lag with 0 for this analysis.

Panel B reports statistics on *RL_WEEKDAY* and the proportion of alerts released within 24 hours after filing for groups formed by the number of news categories in each alert.

based on firm characteristics that capture market demand for information and key firm-specific events likely to heighten market's attention on those firms. We then test our predictions using information on Dow Jones alerts.

Predictions

We posit that market demand for information relates to firm characteristics that capture investor awareness and information environment. Prior studies (Wurgler and Zhuravskaya 2002; Chen et al. 2004; Chen et al. 2006; Docking and Dowen 2006; Elliott et al. 2006; Mase 2006) suggest that firms whose shares are included in a major market index have increased investor awareness. Consistent with this argument, D'Souza et al. (2010b) find that data aggregators place

higher collection priority on the financial statement information of these firms. Therefore, we predict that newswire services are more likely to target periodic reports of firms whose stocks are included in the S&P 500, S&P MidCap 400, or S&P SmallCap index (*SP1500*). All explanatory variables are defined in Table 4.

In addition, we expect that the nature of the information environment to be associated with the demand for newswire services. Extant research suggests that analyst following and institutional ownership are positively associated with the corporate information environment (O'Brien and Bhushan 1990; Frankel et al. 1999; Jiambalvo et al. 2002; Bushman et al. 2004; Piotroski and Roulstone 2004), but each may have a differential information search focus. As analysts have a competitive advantage in extracting industry-level information from firm performance (Piotroski and Roulstone 2004), we predict that newswire alerts could act as a complementary source by providing idiosyncratic information in firms with greater analyst following (*Analysts*). On the contrary, institutional investors are likely to expend greater resources in gathering firm-specific information (Jiambalvo et al. 2002; Piotroski and Roulstone 2004), which could result in newswire services competing with institutions' private search efforts. We therefore refrain from a directional prediction for institutional ownership (*Instown*). We consider the availability of a credit rating (*Rating*) as a proxy for creditor demand, thereby predicting a positive association with the existence of newswire alerts.

Academic research also suggests that when firms experience losses, market participants face an information environment in which traditional profitability measures may be lacking for valuation purposes (Hayn 1995; Collins et al. 1997; Chen et al. 2002; Ajinkya et al. 2005). As additional disclosures of loss firms may be relevant to the marketplace, we also predict that newswires would more often generate alerts for loss firms (*Loss*) to enable investors to better evaluate their financial circumstances.

Li and Ramesh (2009a) find that that a segment of firms first releases its quarterly earnings information only with the periodic reports (also see Amir and Livnat 2005). While the lack of a preliminary earnings release is consistent with lower value relevance of earnings and limited coverage by sophisticated market participants, periodic reports of these firms are more apt to include information hitherto not disclosed to the marketplace. Therefore, we expect newswire services to closely monitor the periodic reports of firms that did not provide a preliminary earnings release (*NoPrelim*). Finally, we consider exposure to litigation risk (*Litigation*) as another factor that is likely to influence newswire services to search for market-moving information in periodic reports.

We next identify key firm-specific events that create heightened market attention in general or point to the existence of potentially value-relevant information in periodic reports. With respect to general market attention, we expect that firms raising capital (*Stockissue* and *Debtissue*), that are takeover targets (*Takeover*), or confronting financial distress (*Ch11* and *Delisting*) face increased market scrutiny and, therefore, would be fitting targets of newswire services, as their periodic reports could contain useful information to interested stakeholders.

With respect to potentially value-relevant information, we first consider circumstances surrounding preliminary earnings releases that would enhance the confirmatory role of periodic reports (Beyer et al. 2009). Specifically, we argue that the market is more likely to demand confirming or negating information in periodic reports when preliminary earnings announcements generate large price reactions (*EA_mktr*) or when firms just meet or beat earnings targets (*JMOB*). In addition, we posit that firms reporting extraordinary items (*Extraordinary*) or special items (*Special*) and firms receiving a nonstandard audit opinion (*Nonstandard*) are likely to include in their periodic filings important information about the economic circumstances they face. Consequently, the periodic reports of such firms are likely to attract news alerts.

To test these predictions, we employ the following empirical model (firm and quarter subscripts are suppressed):

$$prob(Alert) = \Phi\Bigg(\beta_0 + \beta_1 SP1500 + \beta_2 Analysts + \beta_3 Instown + \beta_4 Rating + \beta_5 Loss \\ + \beta_6 NoPrelim + \beta_7 Litigation + \beta_8 Stockissue + \beta_9 Debtissue + \beta_{10} Takeover \\ + \beta_{11} Ch11 + \beta_{12} Delisting + \beta_{13} ER\_mktr + \beta_{14} JMOB + \beta_{15} Extraordinary \\ + \beta_{16} Special + \beta_{17} Nonstandard + \beta_{18} ROA + \beta_{19} Leverage + \beta_{20} Liquidity \\ + \beta_{21} Arbrisk + \beta_{22} Volume + \beta_{23} nEA100 + \beta_{24} nFL100 + \sum_{T=1998}^{2004} \beta_{25,T} YT + \varepsilon \Bigg). \tag{1}$$

As additional controls, we include three performance measures (*ROA*, *Leverage*, and *Liquidity*), given the documented relationship between firm performance and propensity for disclosures (Frankel et al. 1999). Moreover, we add controls for idiosyncratic volatility (*Arbrisk*) and trading volume (*Volume*), given that trading frictions can dampen the demand for corporate accounting information even among sophisticated investors (Collins et al. 2003; Ali et al. 2003; Mashruwala et al. 2006; D'Souza et al. 2010b). Finally, we include proxies for the overall market incidence of earnings announcements (*nEA100*) and periodic reports (*nFL100*) to capture supply-side constraints of newswire services when the market is inundated with corporate accounting information. D'Souza et al. (2010b) document similar effects for a data aggregator.

Empirical Analysis Based on Dow Jones Alerts

Our empirical analysis is based on a sample of 210,620 firm-quarters with a 10-K or 10-Q report, with Dow Jones issuing an alert in less than 9 percent of the cases (18,599/210,620). The low incidence is not surprising, given that the market focuses more on earnings announcements, which increasingly provide detailed financial statement information, and thereby preempt disclosures in periodic filings. Consequently, information intermediaries must have cost-effective screening mechanisms to identify value-relevant information buried in periodic reports.

Our untabulated univariate test results are largely consistent with our predictions. For instance, in support of the investor awareness argument, roughly 38 percent of the alert sample pertains to firms included in one of the S&P major market indices, compared to only about 20 percent in the non-alert sample. Similarly, the median analyst following (institutional ownership) in the alert sample is thrice (more than twice) that of the non-alert sample, consistent with equity holders' demand for information. The evidence on credit demand is strong, with half of the alert sample having an S&P credit rating, as compared to less than a quarter in the non-alert sample. Consistent with nontrivial supply-side effects, periodic reports that do not receive alerts are likely to be filed on days when the median number of filings made with the SEC is 542, as compared to 186 for periodic reports that received alerts.

Turning to the multiple regression analysis, model (1) we estimate using Chamberlain's Random Effects probit estimator, with the results presented in Table 4, Panel A.[14] In addition to reporting marginal effects (measured using partial average effects) and their p-values, we also

---

[14] Two observations are in order. First, Chamberlain's Random Effects probit offers a consistent approach to incorporating unobserved firm-specific heterogeneity as well as the ability to estimate marginal effects. We do not consider fixed

report the economic significance of each independent variable as the marginal effect itself for the indicator variables and the marginal effect times the inter-quartile range for all other explanatory variables.

The overall model is statistically significant, as indicated by the Wald Chi-squared statistic, with a Pseudo R^2 of 13.5 percent. Thirteen of the 16 predicted effects have statistically significant marginal effects with the correct signs at the two-tailed 0.10 level. Given that the unconditional probability of an alert from Dow Jones is around 0.09, the reported economic significance of many of the predicted effects appears material. We list below variables for which the economic significance is at least half of a percentage point and the statistical significance is at least at the 0.05 level:

Variable	***Nonstandard***	***NoPrelim***	***Rating***	***Delisting***	***SP1500***	***Litigation***	***Special***	***Loss***	***Stockissue***
Economic Significance	0.051	0.037	0.030	0.019	0.012	0.011	0.007	0.005	0.005

Taken together, we find that Dow Jones staff considers investor awareness, information environment, and key firm-specific events to identify potentially value-relevant information from periodic SEC reports. Of the control variables, the supply-side proxies show the largest effect on the incidence of alerts, with the probability of alerts declining by 0.017 when the intensity of earnings releases in the marketplace (*nEA100)* increases by its inter-quartile range (221). The evidence is consistent with the perceived priority of earnings announcements and, consequently, the diversion of Dow Jones' resources away from the periodic reports.

We next provide descriptive evidence on factors associated with the speed with which Dow Jones releases the alert.[15] In Table 4, Panel B we report the results of a pooled Poisson regression on the subsample with alerts in which the dependent variable is the alert lag (*RL_WEEKDAY*) and the independent variables are as in model (1).[16] Given that we use a Poisson regression, the economic significance levels we report are percentage changes in alert lag for a unit increment for all indicator variables and at inter-quartile range for all continuous variables (Wooldridge 2003, 574). Conditional on issuing an alert, we find that only a few of the variables are statistically significant in explaining the alert lag. However, we find some effects with large economic significance in that the alert lags are roughly 40 (22) percent shorter for firm-quarters without a preliminary earnings announcement (for firms with higher litigation risk). In addition, consistent with supply-side constraints, the alert lag is affected by information overload, with about a 9 (16) percent delay for an inter-quartile range increment in the number of earnings announcements (periodic SEC filings) released on the same day as the alerted filings.

---

effects probit or fixed effects logit, as the former is inconsistent and the latter does not identify marginal effects (Wooldridge 2002, Chapter 15). Second, from an implementation standpoint, the Chamberlain's Random Effects approach models unobserved heterogeneity as a linear function of firm-specific means of all time-varying independent variables in the panel data plus an error term with zero mean and variance s^2_e. Consequently, the Chamberlain's Random Effects probit is a pooled regression that includes the firm-specific means of all time-varying independent variables as additional explanatory variables to control for unobserved heterogeneity at the firm level.

15 Periodically, Dow Jones sends summary reports to its clients touting its ability to beat its competition in identifying market-moving information from SEC filings. For instance, in its March 2009 issue of Dow Jones Financial Services Solutions' *The Edge*, Dow Jones states that "on Feb. 16, the Friday ahead of a three-day weekend for the U.S., the Dow Jones Corporate Filings Alert scooped the competition with its report, unmatched by our competitors until four days later, that State Street Corp. disclosed billions of dollars in unrealized losses in its investment portfolio and the off-balance-sheet entities it manages."

16 We do not report fixed effects Poisson results because the time-invariant indicator *Litigation*, which has significant influence on the alert lag, would be omitted from the regression. However, the inferences of fixed effects Poisson are very similar to those of pooled Poisson, except that *EA_mktr* and *Nonstandard* become insignificant.

TABLE 4

Determinants of the Issuance and Timing of Corporate Filing Alerts

Panel A: Determinants of Corporate Filing Alert

	Pred. Sign	Marginal Effect	p-value	Economic Sig.
SP1500	+	0.0116	(0.019)**	0.0116
Analysts	+	0.0006	(0.065)*	0.0031
Instown	+/−	−0.0004	(0.000)***	−0.0195
Rating	+	0.0299	(0.000)***	0.0299
Loss	+	0.0054	(0.017)**	0.0054
NoPrelim	+	0.0370	(0.000)***	0.0370
Litigation	+	0.0107	(0.000)***	0.0107
Stockissue	+	0.0053	(0.020)**	0.0053
Debtissue	+	−0.0013	(0.601)	−0.0013
Takeover	+	0.0082	(0.105)	0.0082
Ch11	+	−0.0141	(0.097)*	−0.0141
Delisting	+	0.0192	(0.001)***	0.0192
EA_mktr	+	0.0578	(0.002)***	0.0020
JMOB	+	−0.0005	(0.774)	−0.0005
Extraordinary	+	0.0067	(0.061)*	0.0067
Special	+	0.0071	(0.000)***	0.0071
Nonstandard	+	0.0512	(0.000)***	0.0512
ROA		−0.0582	(0.000)***	−0.0017
Leverage		0.0127	(0.078)*	0.0040
Liquidity		−0.0181	(0.051)*	−0.0078
Arbrisk		0.0012	(0.008)***	0.0035
Volume		0.0002	(0.000)***	0.0008
nEA100		−0.0078	(0.000)***	−0.0171
nFL100		−0.0007	(0.000)***	−0.0092
Year dummies?			Yes	
Obs.			147,119	
Pseudo R^2			0.135	
Wald Chi-squared			5,806.67	

Panel B: Determinants of Alert Lag (*RL_WEEKDAY*)

	Pred. Sign	Marginal Effect	p-value	Economic Sig.
SP1500	−	0.0092	(0.790)	0.0092
Analysts	−	−0.0035	(0.211)	−0.0177
Instown	−/+	−0.0003	(0.626)	−0.0128
Rating	−	0.0146	(0.662)	0.0146
Loss	−	−0.0233	(0.417)	−0.0233
NoPrelim	−	−0.3979	(0.000)***	−0.3979
Litigation	−	−0.2237	(0.000)***	−0.2237
Stockissue	−	−0.0023	(0.913)	−0.0023
Debtissue	−	0.0494	(0.242)	0.0494
Takeover	−	0.0202	(0.744)	0.0202
Ch11	−	−0.0316	(0.808)	−0.0316
Delisting	−	−0.0222	(0.749)	−0.0222

(continued on next page)

Panel B: Determinants of Alert Lag (*RL_WEEKDAY*)

	Pred. Sign	Marginal Effect	p-value	Economic Sig.
EA_mktr	−	−0.8969	(0.005)***	−0.0310
JMOB	−	−0.0248	(0.321)	−0.0248
Extraordinary	−	−0.0407	(0.314)	−0.0407
Special	−	−0.0106	(0.689)	−0.0106
Nonstandard	−	0.0969	(0.006)***	0.0969
ROA		0.0611	(0.691)	0.0018
Leverage		−0.0516	(0.366)	−0.0165
Liquidity		0.0625	(0.391)	0.0269
Arbrisk		−0.0071	(0.283)	−0.0208
Volume		−0.0017	(0.000)***	−0.0074
nEA100		0.0401	(0.000)***	0.0885
nFL100		0.0124	(0.000)***	0.1627
Year dummies?			Yes	
Obs.			14,940	
Pseudo R^2			0.096	
Wald Chi-squared			738.25	

*, **, *** Represent statistical significance at two-tailed 0.1, 0.05, and 0.01 levels, respectively.

Panel A reports the Chamberlain's Random Effects probit regression results of factors that trigger the issuance of Dow Jones corporate filing alerts, while Panel B reports the pooled Poisson regression results of factors that determine the alert lag between the periodic SEC report filing time and the release time of the initial alert, measured in weekdays. Given that the estimated Chamberlain's Random Effects coefficients are not directly interpretable, we report the marginal effects (calculated as the partial average effect) and their two-tailed p-values based on firm-clustered standard errors. The marginal effects in Panel B equal the estimated Poisson coefficients. The economic significance is the same as the marginal effect for indicators and calculated as the marginal effect multiplied by the inter-quartile range for continuous variables.

Variable Definitions:

- *SP1500* = 1 if the firm is in the S&P 1500 index at the end of current fiscal quarter, and 0 otherwise;
- *Analysts* = number of analyst following over the current fiscal quarter; we obtain analyst following from the I/B/E/S database;
- *Instown* = percentage of institutional stock ownership (winsorized at 100) at the end of the calendar quarter at or immediately preceding the current fiscal quarter; we obtain institutional ownership from Thomson Financial Spectrum;
- *Rating* = 1 if S&P credit rating is available for the firm in the current fiscal quarter, and 0 otherwise;
- *Loss* = 1 if net income before extraordinary items is negative for the current fiscal quarter, and 0 otherwise;
- *NoPrelim* = 1 if the earnings information is first released through periodic SEC reports for the current fiscal quarter, and 0 otherwise. This includes cases where there is no preliminary earnings announcement, earnings announcement date concurs with the periodic SEC filing date, or the earnings announcement is made after the periodic SEC report is filed;
- *Litigation* = 1 for industries found to be exposed to high litigation risk: biotech (SIC 2833–2936), computer hardware (SIC 3570–3577), electronics (SIC 3600–3674), retail (SIC 5200–5961), and computer software (SIC 7371–7379), and 0 otherwise;
- *Stockissue* = 1 for any of the six quarters leading to the firm's stock issuance (including the quarter of stock issuance) and 0 otherwise; we obtain the stock issue information from *SDC*;
- *Debtissue* = 1 for any of the six quarters leading to the firm's public debt issuance (including the fiscal quarter of debt issuance), and 0 otherwise; we obtain the public debt issue information from *SDC*;
- *Takeover* = 1 for any of the six quarters leading to the firm's delisting from a stock exchange because of mergers and acquisitions (including the quarter of takeover), and 0 otherwise;
- *Ch11* = 1 for quarters [−5, 5] centered on the fiscal quarter when the firm files Chapter 11 petition, and 0 otherwise; we obtain the Chapter 11 bankruptcy information from Professor Lynn M. LoPucki's Bankruptcy Research Database;
- *Delisting* = 1 for any of the six quarters leading to the firm's delisting from a stock exchange because of financial distress (including the quarter of delisting), and 0 otherwise;

(continued on next page)

EA_mktr = market reaction at the earnings announcement day, measured as absolute value of daily return minus value-weighted market return and winsorized at its top and bottom 0.5 percentile values; this variable is set to 0 if *NoPrelim* equals 1;
JMOB = 1 if the earnings surprise is between \$0.00 and \$0.03 per share, and 0 otherwise. Earnings surprise is measured as actual earnings per share (EPS) minus median of the latest five individual EPS forecasts made over a period from the previous quarter's earnings announcement to the current quarter's earnings announcement. Both actual value and forecasts of EPS are obtained from the unadjusted I/B/E/S database. If I/B/E/S data are not available for a firm-quarter, then we define earnings surprise as seasonally adjusted EPS (i.e., EPS in the current quarter minus EPS in the same quarter of prior year);
Extraordinary = 1 if an extraordinary item is reported in the Compustat database for the current fiscal quarter q, and 0 otherwise;
Special = 1 if a special item is reported in the Compustat database for the current fiscal quarter q, and 0 otherwise;
Nonstandard = 1 if the auditor's opinion on the firm's financial statements for the current fiscal year is not "unqualified," and 0 otherwise; we obtain auditor's opinion from the Compustat database (item #149) and this variable is set to 0 for interim quarters;
ROA = net income before extraordinary items divided by total assets at the beginning of the current fiscal quarter, winsorized at its top and bottom 0.5 percentile values;
Leverage = long-term debt divided by total assets at the end of the current fiscal quarter, winsorized at its top and bottom 0.5 percentile values;
Liquidity = sum of cash and cash equivalents and receivables divided by total assets at the end of the current fiscal quarter, winsorized at its top and bottom 0.5 percentile values;
Arbrisk = standard deviation of residuals from a regression of firm-specific daily returns on the returns of the CRSP equally weighted market index over the current fiscal quarter (a minimum of five observations is required); we multiply the standard deviation of residuals by 100 and winsorize it at its top and bottom 0.5 percentile values;
Volume = daily trading volume (in millions of dollars) averaged over the current fiscal quarter, winsorized at its top and bottom 0.5 percentile values;
nEA100 = number of all earnings announcements (in 100s) made on the day of the firm's filing of its periodic SEC report of the current fiscal quarter; and
nFL100 = number of periodic SEC filings (in 100s) released on the day of the firm's filing of its periodic SEC report of the current fiscal quarter.

V. DOES THE MARKET REACT TO NEWSWIRE ALERTS ON PERIODIC REPORTS?

In this section, we first use daily returns and trading volume to examine the market reaction to the alerts issued by Dow Jones. For those alerts issued during trading hours, we next examine intra-day price and volume reactions to isolate the immediate trading implications of the alerts and control for any endogeneity present in the analysis based on daily data.

Daily Returns and Volume Analysis

Using daily data over the period 1997–2004, we develop a multiple regression model to jointly examine market reactions to Dow Jones alerts after controlling for the effects of periodic filings and earnings announcements:

$$DV_t = \alpha + \sum_{Event=1}^{7} \sum_{Day=-2}^{2} \beta^{Event,Day} \cdot Indicator_t^{Event,Day}, \tag{2}$$

where DV is a measure of daily stock price or volume reaction (defined below). The independent variables are 35 indicator variables corresponding to five event days [−2 to +2] surrounding each of the following seven events:

- Dow Jones alerts;
- Two types of earnings announcements (interim, annual);

- Two types of 10-K reports (without an alert, followed by an alert); and
- Two types of 10-Q reports (without an alert, followed by an alert).[17]

We include separate indicator variables for periodic reports with and without a Dow Jones alert to control for any characteristics of the alert firm-quarters that might engender a differential reaction to the periodic filing.

Following Li and Ramesh (2009a), the dependent variable for stock returns is based on the absolute value of the difference between daily return and value-weighted market return. Following Cready and Mynatt (1991), the trading volume measure is based on the natural logarithm of the sum of volume and 0.000255, where volume equals daily trading volume divided by total outstanding shares. We standardize both the stock returns and trading volume measures by subtracting their firm-year mean and dividing by their firm-year standard deviation and denote them as *SAER* and *STV*, respectively.[18] Table 5, Panel A indicates that, after controlling for earnings announcements and periodic SEC filings, firm-quarters with a Dow Jones alert elicit statistically significant price and volume reactions on the event day (day 0).[19]

In Table 5, Panel B, we report the results of an expanded regression in which we break down the alert days into three sub-groups: (1) the initial alert days with a single alert; (2) the initial alert days containing multiple alerts; and (3) issuance of follow-up alerts subsequent to the initial alert day (coefficients of all other events are suppressed). The results indicate that the market reactions are limited to the initial alert day regardless of whether it is followed by another day with an alert. While the follow-up alerts may have market implications, our results based on daily data do not provide any evidence of their market-moving effects.

While this study and the extant research find no immediate price or volume reactions to the filing of 10-Q reports (Li and Ramesh 2009a; Lu 2006), one unanswered question is whether alerts based on 10-Q reports generate a market reaction, which we examine next. Given that the market reaction to follow-up alert days is not statistically significant, we focus on the effect of all initial alert days combined when comparing between 10-K and 10-Q alerts. The results in Table 4, Panel C show that both 10-K- and 10-Q-based alerts are associated with significant price and volume reactions (all other coefficients suppressed). Taken together, we find that initial alerts, whether based on annual or quarterly reports, provide incrementally value-relevant information.

We next provide evidence on the differential value relevance of various news categories. Given the general market reactions to all alerts, we estimate the following conditional regression based on 17,454 *Alert Day* observations with non-missing price and trading volume data:

$$\begin{aligned} DV_t = {} & \beta_0 + \beta_1 PER_t + \beta_2 CRE_t + \beta_3 EQU_t + \beta_4 BSC_t + \beta_5 LGL_t + \beta_6 EMP_t + \beta_7 CBS_t + \beta_8 FOR_t \\ & + \beta_9 BKC_t + \beta_{10} TAX_t + \beta_{11} OTH_t + \beta_{12} Missing_t + \beta_{13} Unclassified_t + \beta_{14} Multiple_t \\ & + \beta_{15} Second\ Alert\ Day_t + \beta_{16} Third\ and\ Following\ Alert\ Days_t + \varepsilon_t. \end{aligned} \tag{3}$$

Regression (3) includes indicators for each of the news categories as well as controls for alerts with subject codes missing (*Missing*) or not classifiable as firm-specific news (*Unclassified*). Given that the news categories are not mutually exclusive, we are able to include indicator variables for each of them. In addition, we include an indicator for days with multiple alerts

[17] If the alerts or periodic reports are released after the market closes (i.e., post 4:00 p.m. Eastern time), we re-center the five-day event window around the next business day.

[18] In untabulated analysis, we use a Z-transformation of rank scores (Blom 1958) of absolute excess returns and share turnover as nonparametric measures for our dependent variables (Easton and Zmijewski 1993; Carter and Soo 1999; Li and Ramesh 2009a). The tenor of our market reaction results remains unaltered.

[19] See Li and Ramesh (2009a) for a discussion of a calendar quarter-end effect surrounding the filing of 10-K reports as well as evidence of a possible information transfer during quarter-ends. Our inferences regarding the market reaction to alerts are identical after controlling for the quarter-end effect (untabulated).

TABLE 5

Regression Results of Market Reaction Surrounding Corporate Filing Alerts

Panel A: OLS Regression Results of Market Reactions to All Alerts

		Standardized Absolute Excess Return (*SAER*)				
	Pred. Sign	**Day −2**	**Day −1**	**Day 0**	**Day 1**	**Day 2**
Event 1: Annual Earnings Release (n = 43,194)	+	0.027***	0.076***	0.433***	0.506***	0.125***
Event 2: Interim Earnings Release (n = 131,185)	+	0.043***	0.109***	0.473***	0.518***	0.127***
Event 3: 10-Ks without Alerts (n = 37,696)	+	−0.013	0.005	0.054***	0.046***	0.070***
Event 4: 10-Ks with Alerts (n = 5, 600)	+	−0.027	0.012	0.099***	0.055***	0.040***
Event 5: 10-Qs without Alerts (n = 120,668)	+	−0.063	−0.073	−0.057	−0.080	−0.054
Event 6: 10-Qs with Alerts (n = 10,659)	+	−0.060	−0.064	−0.016	−0.071	−0.056
Event 7: All Alerts (n = 17,561)	+	0.009	−0.023	0.033***	−0.011	−0.003

		Standardized Trading Volume (*STV*)				
	Pred. Sign	**Day −2**	**Day −1**	**Day 0**	**Day 1**	**Day 2**
Event 1: Annual Earnings Release (n = 43,194)	+	−0.020***	0.081***	0.503***	0.625***	0.341***
Event 2: Interim Earnings Release (n = 131,185)	+	−0.031***	0.064***	0.457***	0.564***	0.287***
Event 3: 10-Ks without Alerts (n = 37,696)	+	0.004	0.011**	0.049***	0.023***	0.034***
Event 4: 10-Ks with Alerts (n = 5,600)	+	−0.004	−0.005	0.037***	0.031**	0.047***
Event 5: 10-Qs without Alerts (n = 120,668)	+	−0.051	−0.066	−0.063	−0.083	−0.058

(continued on next page)

	Pred. Sign	Standardized Trading Volume (*STV*) Day −2	Day −1	Day 0	Day 1	Day 2
Event 6: 10-Qs with Alerts (n = 10,659)	+	−0.036	−0.035	−0.009	−0.060	−0.029
Event 7: All Alerts (n = 17,561)	+	−0.006	−0.032	0.029***	−0.014	−0.016

Panel B: OLS Regression Results of Market Reactions to Initial Alerts versus Follow-Up Alerts

	Pred. Sign	Standardized Absolute Excess Return (*SAER*) Day −2	Day −1	Day 0	Day 1	Day 2
Initial - Single Alerts (n = 14,453)	+	0.011	−0.029	0.033***	−0.016	−0.004
Initial - Multiple Alerts (n = 1,799)	+	0.012	0.039*	0.056**	0.023	0.007
Follow-up Alerts (n = 1,309)	+	−0.005	−0.023	0.014	0.006	−0.006

	Pred. Sign	Standardized Trading Volume (*STV*) Day −2	Day −1	Day 0	Day 1	Day 2
Initial - Single Alerts (n = 14,453)	+	0.006	−0.021	0.044***	0.002	−0.004
Initial - Multiple Alerts (n = 1,799)	+	−0.037	−0.010	0.057**	0.011	−0.004
Follow-up Alerts (n = 1,309)	+	−0.024	−0.070	−0.043	−0.081	−0.049

Panel C: OLS Regression Results of Market Reactions to 10-K Alerts versus 10-Q Alerts

	Pred. Sign	Standardized Absolute Excess Return (*SAER*) Day −2	Day −1	Day 0	Day 1	Day 2
Alerts following 10-Ks (n = 5,598)	+	0.015	−0.040	0.043***	0.016	0.022*
Alerts following 10-Qs (n = 10,654)	+	0.008	−0.012	0.032***	−0.026	−0.015

(continued on next page)

	Pred. Sign	Standardized Trading Volume (*STV*) Day −2	Day −1	Day 0	Day 1	Day 2
Alerts following 10-Ks (n = 5,598)	+	−0.001	−0.031	0.046***	0.018	0.027**
Alerts following 10-Qs (n = 10,654)	+	0.002	−0.014	0.045***	−0.005	−0.021

*, **, *** Represent statistical significance at one-tailed 0.1, 0.05, and 0.01 levels, respectively.

This table reports OLS regression results of market reactions based on a sample of 12,201,115 daily observations between 1997 and 2004. *SAER* is calculated as absolute value of daily return minus value-weighted market return, while *STV* is calculated as log (volume + 0.000255), where volume equals to daily trading volume divided by outstanding shares. We normalize both measures by subtracting firm/year mean and then dividing by firm/year standard deviation. In the Panel A regressions, we include 35 indicators for event days [−2, +2] surrounding the following events: (1) Dow Jones alerts; (2) two types of earnings announcements (interim and annual); (3) two types of 10-K reports (without an alert, followed by an alert); and (4) two types of 10-Q reports (without an alert, followed by an alert). If the alerts or periodic reports are released after the market closes (i.e., post 16:00 Eastern time), we re-center the five-day event window to the next business day. In the Panel B regressions, we break down alert event days into three groups: initial alert days with a single alert; initial alert days containing multiple alerts; and issuance of follow-up alerts subsequent to the initial alert day. In the Panel C regressions, we further break down all initial alert days into 10-K- and 10-Q-based alerts. For brevity, we omit the coefficients for all other indicators in Panel B and Panel C as the inferences remain the same. The tenor of our results is similar when we use time-clustered standard errors to adjust for cross-sectional dependence.

(*Multiple*), and indicators for the second trading day with another alert(s) based on the same periodic report (*Second Alert Day*), and alert days beyond the second day for the same periodic report (*Third and Following Alert Days*).

The statistically significant return and volume model intercepts in Table 6 reflect the average market reactions reported in Table 5 for the alert subsample. When discussing Table 6 results, we limit our discussions to indicator variables with consistent price and volume reactions. Compared with other news categories, the legal- and bankruptcy-related alerts generate both higher price reaction and trading volume, whereas the alerts with missing subject codes generate significantly lower market reactions. The results for bankruptcy-related alerts are consistent with the deterioration in the conventional public sources of information on distressed companies coupled with the demand from sophisticated market participants for trading in such companies (Gilson 1995).[20]

The lower price and volume reactions to alerts without a specific news category suggest that immediate trading based either on algorithms or requiring human intervention is constrained in these circumstances. As in the expanded regression results, we find that follow-up alerts issued one or more days after the initial alert generate much less market activity. The lower market reaction for delayed alerts is consistent with the limits to any single information intermediary's ability to consistently identify market-moving information hidden in periodic reports long after their filing. Overall, while the market responds most strongly to legal- and bankruptcy-related alerts, alerts without a specific news category and delayed alerts may not be value-relevant.

Finally, we compare the results in Tables 4 and 6 to examine whether factors that influence Dow Jones' alert decisions also result in significant market reaction when alerts contain news relating to these factors. Given that the analysis in Table 6 is based on the conditional sample of alerts, the reported coefficients merely capture the incremental market reaction. To measure the total market reaction, we report the sum of the coefficient corresponding to each of the news categories and the intercept (Coeff. + Const.) in Table 6.[21] In Table 4 we find that Dow Jones is more likely to issue alerts for firms with a credit rating or when they issue equity. Consistent with this evidence, we find that equity-related (*EQU*) and credit-related (*CRE*) news categories generate significant return and volume reactions.

Intra-Day Returns and Volume Analysis

One obvious concern with using daily data is the possible endogeneity that market participants are more focused on firm-quarters with an alert for reasons other than the information contained in the alerts. We partially control for this possibility in Table 5 by including separate indicator variables for periodic reports with and without a Dow Jones alert in addition to indicators for the alerts themselves. In this subsection we attempt to isolate the price and volume reactions to a narrow intra-day period surrounding the release of Dow Jones alerts, thereby potentially eliminating other confounding effects.

We start with the subsample of alerts issued within trading hours (15,403), and then limit our focus to those alerts issued between 10:30 a.m. and 3:00 p.m. (11,059) to allow for the 60-minute

---

[20] Consistent with our result for the legal-related alerts, our discussions with a money manager indicate that some institutional investors rely on information intermediaries to track litigation-related disclosures or information of companies they follow.

[21] Coeff. + Const. captures the total market reaction to an initial single alert that contains news content corresponding to the specific category (i.e., setting *Multiple*, *Second Alert Day*, and *Third and Following Alert Days* to zero).

TABLE 6
Regression Results of Market Reaction by News Categories

	SAER		*STV*	
	Coeff.	**Coeff. + Const.**	**Coeff.**	**Coeff. + Const.**
Performance (*PER*)	−0.0405	0.1236	0.0026	0.0909
	(0.195)	(0.000)***	(0.926)	(0.000)***
Credit Related (*CRE*)	−0.0449	0.1193	−0.0196	0.0687
	(0.021)**	(0.000)***	(0.277)	(0.028)**
Equity Related (*EQU*)	−0.0471	0.1170	−0.0292	0.0591
	(0.028)**	(0.001)***	(0.152)	(0.080)*
Business Structure Change (*BSC*)	−0.0714	0.0928	−0.0133	0.0750
	(0.004)***	(0.012)**	(0.589)	(0.036)**
Corporate Legal Issue (*LGL*)	0.0558	0.2199	0.0849	0.1732
	(0.079)*	(0.000)***	(0.002)***	(0.000)***
Employment Related (*EMP*)	0.0257	0.1898	0.1063	0.1946
	(0.598)	(0.001)***	(0.022)**	(0.000)***
Corporate Business (*BUS*)	0.0143	0.1785	0.0924	0.1807
	(0.800)	(0.003)***	(0.077)*	(0.002)***
Forecast (*FOR*)	−0.0531	0.1111	−0.0055	0.0827
	(0.049)**	(0.004)***	(0.830)	(0.021)**
Bankruptcy (*BKC*)	0.1403	0.3045	0.2341	0.3224
	(0.056)*	(0.000)***	(0.001)***	(0.000)***
Tax Related (*TAX*)	0.0117	0.1759	0.0361	0.1244
	(0.883)	(0.043)**	(0.629)	(0.127)
Other (*OTH*)	0.0383	0.2024	0.0848	0.1730
	(0.294)	(0.000)***	(0.004)***	(0.000)***
Missing	−0.3026	−0.1384	−0.2693	−0.1810
	(0.000)***	(0.055)*	(0.001)***	(0.014)**
Unclassified	0.0693	0.2335	0.0816	0.1699
	(0.448)	(0.006)***	(0.265)	(0.011)**
Multiple	0.0576		0.0386	
	(0.039)**		(0.141)	
Second Alert Day	−0.0821		−0.0991	
	(0.012)**		(0.002)***	
Third and Following Alert Days	−0.0861		−0.2143	
	(0.218)		(0.003)***	
Constant	0.1641		0.0883	
	(0.000)***		(0.004)***	
R^2	0.0034		0.0060	

*, **, *** Represent statistical significance at two-tailed 0.1, 0.05, and 0.01 levels, respectively.

This table reports market reaction results by news categories based on 17,343 Alert Days with market reaction data. *SAER* and *STV* are as defined in Table 5. We create indicators for each news category as well as controls for alerts with subject codes missing (*Missing*) or not classifiable as firm-specific news (*Unclassified*). We cannot categorize 276 alerts whose subject codes do not indicate any firm-specific news categories (e.g., "Corporate/Industrial News," "Commodity/Financial Market News"). *Multiple* equals 1 if multiple alerts for the same filing are released in one day, and 0 otherwise. *Second Alert Day* (*Third and Following Alert Days*) equals 1 if it is the second alert day (third and following alert days) after the initial alert day, and 0 otherwise. For each news category indicator in the regressions, we report both estimated coefficient and the sum of the coefficient and the intercept. While the former describes the differential value relevance of each news category, the latter captures the total market reaction to an initial single alert that contains news content corresponding to the specific news category (i.e., setting *Multiple*, *Second Alert Day*, and *Third and Following Alert Days* to 0). Two-tailed p-values are calculated based on firm-clustered standard errors and presented in parentheses.

pre- and post-event windows that we examine.[22] After merging with the TAQ database, we have 7,913 alerts with complete trading data. In the spirit of Brown et al. (1992, 77–78), corresponding to each alert we identify a pseudo-event (i.e., a pseudo alert) as occurring at the same hour and minute as the alert and on the same weekday of the week that immediately follows the alert. We choose the pseudo-events after the issuance of the alerts so that earnings announcements are less likely to confound our inferences.

For both pseudo and actual alerts, our intra-day return and volume metrics are measured using the TAQ trade files by keeping only trades that meet all of the following criteria (Ng et al. 2008): (1) trades occurred on the NYSE, AMEX, or NASDAQ; (2) trades were made under regular market conditions; (3) trades were made within the normal trading hours (i.e., 9:30 a.m.–4:00 p.m.); (4) trades were good trades without subsequent cancellations; and (5) the transaction price and the number of shares traded were both positive. We consider a window starting 60 minutes prior to the event minute to 60 minutes following (a total of 121 minutes). For each minute in the event window, we calculate the stock return measure (*ABSRET*) as the absolute value of $\frac{(PRICE_t - PRICE_{t-1})}{PRICE_{t-1}}$, where $PRICE_t$ is the trading price of the last transaction within minute t and $PRICE_{t-1}$ is the trading price of the last transaction before minute t. If no transaction occurs within minute t, then *ABSRET* is set to 0. We use the number of shares traded within the minute deflated by shares outstanding as the metric for trading volume, labeled *SVOL*. Untabulated analysis using the un-deflated number of shares traded yields similar results.

In the two panels of Figure 2 we plot the average minute-by-minute *ABSRET* and *SVOL*, respectively, for the Dow Jones alerts versus the pseudo-events. In both cases, we find no noticeable market activity surrounding the pseudo-events, consistent with their proxy for non-event behavior. However, we find almost instantaneous price and volume reactions to the issuance of Dow Jones alerts, with persistent market movements for several minutes following the alert. Compared to the average *ABSRET* of the 60-minute pre-event period, the absolute returns for the first five minutes following the alert (including the minute of the alert) are between 28 to 52 percent higher, with the corresponding figures of −11 to 3 percent for the pseudo-events (untabulated).[23] Results based on *SVOL* are similar, with 18 to 69 percent higher trading activity following the alert, compared to between −4 and 4 percent following the pseudo-events (untabulated).[24] In comparison to the market reaction to the alerts, the untabulated analysis provides no evidence of incremental market movements following the filing of periodic reports that triggered one or more alerts.[25]

---

[22] Two observations are in order. First, limiting our alerts to those issued between 10:30 a.m. and 3:00 p.m. could introduce another selection bias. However, we believe the benefits of our sample restriction outweigh any costs as the opening-of-day and end-of-day market effects documented in the finance literature (Gerety and Mulherin 1992) would inextricably confound our analysis based on the full sample. Moreover, we believe that the incentives of the newswire service should be unaltered by the timing of the alert, so the results based on our restricted sample should have broader implications. Second, Brown et al. (1992) use a similar restriction in their analysis to control for the opening-of-day and end-of-day effects.

[23] Kim et al. (1997) do not find price reactions to the public release of analyst recommendations by Dow Jones through the Broad Tape. The lack of a price reaction is not surprising given that Dow Jones is merely re-releasing information that was privately distributed to important analyst clients before the market opened. In fact, Kim et al. (1997) find that the information released privately to the clients was impounded during the first 5 to 15 minutes after the market opened. In our context, while the periodic reports are publicly available, Dow Jones screens for and identifies market-moving information from them. Indeed, our results suggest that the screening service of Dow Jones has informational value to the marketplace.

[24] We also find that the average number of trades in the first 16 minutes $[0, +15]$ subsequent to an alert is 30 as compared to 24 in the same window following a pseudo alert, representing a 25 percent increase.

[25] In addition, we plot the price and volume activities of the 48,565 SEC filings during the sample period that did not receive a Dow Jones filing alert and are filed between 10:30 a.m. and 3:00 p.m. To address the quarter-end effect documented in Li and Ramesh (2009a), we divide the filings into three groups: 10-K QEND (3,594), 10-K NQEND

We next formally test the significance of the price and volume reactions to the release of Dow Jones alerts. We define the event window as consisting of the event minute corresponding to the alert (or SEC filing) time stamp and the 15 minutes immediately following the time stamp. To avoid confounds, we further limit our analysis to periodic reports and alerts that are more than 15 minutes apart. We define pre- and post-event periods of the same length (denoted $[-16, -1]$ and $[+16, +31]$) for benchmarking. In addition, using the pseudo-events discussed above, we similarly define three event windows. For each of the 16-minute windows, we calculate the absolute return and share turnover measures. The significance of t-tests for the mean absolute returns and share turnover at the alert and SEC filing event windows are reported in Table 7, Panels A and B, respectively. We also report test results separately for quarterly and annual reports and their corresponding alerts.

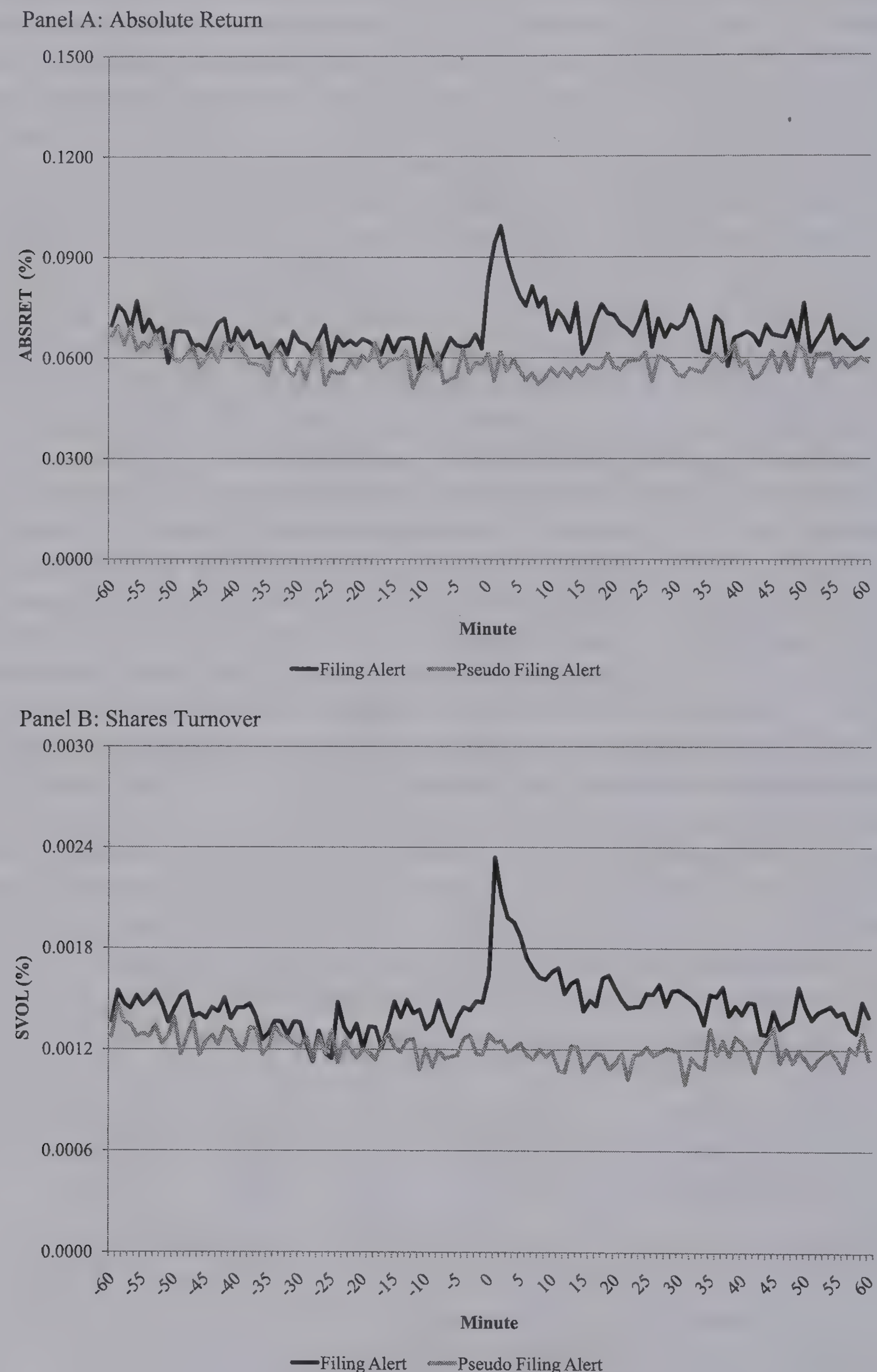
In columns (4) and (5) of both panels, we compare the "true" event-window price or volume measure with that of the pre- and post-event window, respectively. In columns (9) and (10), we report similar test results, but based on pseudo-event windows. Finally, in column (11), we compare the "true" event-window price or volume measure with that of the pseudo-event-window. Focusing on the pseudo-event results for the alerts and the SEC filings in columns (9) and (10), the event-window returns and volume behave in a similar fashion to those in both pre- and post-event window, with only 4 out of 24 comparisons providing significant differences. Overall, the pseudo-event window market reaction appears to be a reasonable benchmark for testing the information content of Dow Jones alerts. In addition, the results are unlikely driven by a small number of trades, in that during the first 16 minutes subsequent to an alert (pseudo alert), 61.5 (59.2) percent of the observations indicate a price change and 75.0 (73.3) percent have nonzero trading volume (untabulated).

As shown in Table 7, Panel A for all alerts ("All"), when compared to the pre- or post-event window, we find a statistically significant increase in market activity at the event window capturing the release of Dow Jones alerts (columns (4) and (5)). This result holds when we benchmark the alert event window against the pseudo-event window (column (11)). Note that the economic significance is larger when we benchmark against the pseudo-event window. For instance, the alert event price reaction is 35 percent larger when compared to the pseudo-event, but it is only 18 percent larger when compared to the pre-event period. The corresponding figures for the trading volume are 47 percent for the pseudo-event comparison and 22 percent for the pre-event comparison.[26] When we break down alerts between those that are 10-K-based versus 10-Q-based, the results are largely consistent, in that both sub-groups show significant market reactions. The results in Panel B indicate that, when compared to the pre- or post-event period, the periodic SEC reports do not generate significantly larger price or volume activity during the 16-minute filing window using any of the three benchmarks.

(6,038), and 10-Q (38,933), where 10-K QEND (NQEND) refers to 10-Ks filed within (outside of) the $[-1, +1]$ trading day window around any calendar quarter-end. 10-Q refers to all 10-Q filings. The price and volume charts for all three groups show no evidence of a market reaction around the filing of the periodic reports. This evidence dispels the concern that the lack of an immediate intra-day market reaction to filings that are followed by alerts may not be generalized to other periodic SEC filings. Consistent with Li and Ramesh (2009a), 10-Q graphs show no noticeable market activity throughout the event window, whereas non-quarter-end 10-K graphs show higher market volatility but similar price and trading levels when compared to the 10-Qs. The quarter-end 10-Ks are associated with the highest volatility and trading/price levels among the three groups, but the high volatility and trading levels are present during the entire 121-minute window, consistent with a quarter-end effect (Li and Ramesh 2009a).

[26] Consistent with our analysis of determinants of alert lag (Table 4, Panel B), we find that the longer it takes for Dow Jones to issue an alert, the lower is the market reaction. One-standard-deviation increment in the number of minutes between the filing time and the alert time reduces the 16-minute stock return (volume) reaction by 15.8 (16.9) percent (untabulated).

FIGURE 2
Average Intra-Day Market Reaction to Dow Jones Filing Alerts, by Minute



These figures are based on 7,913 Dow Jones filing alerts issued between 10:30 a.m. and 3:00 p.m. of trading days over 1997–2004. The corresponding pseudo-event window starts with the same hour and minute as the filing alert and on the same weekday of the week that immediately follows the filing alert. *ABSRET* is calculated as the absolute value of $\frac{(PRICE_t - PRICE_{t-1})}{PRICE_{t-1}}$, where $PRICE_t$ is the trading price of the last transaction within minute t and $PRICE_{t-1}$ is the trading price of the immediately previous transaction before minute t. If no transaction occurs

within minute *t*, then *ABSRET* is set equal to 0. Shares turnover is the number of shares traded within the minute deflated by total shares outstanding (*SVOL*). All measures have been winsorized at their respective top 0.1 percent.

Sensitivity Analyses and Robustness Checks

To corroborate our findings, we conduct a series of untabulated sensitivity analyses and robustness checks. One concern about the immediate market reactions to Dow Jones alerts is that they may merely reflect noise trading as opposed to trading behavior consistent with improved price efficiency. If so, one would expect complete price reversals as informed traders enter to correct any mispricing or over-reaction. When we classify Dow Jones alerts into good and bad news based on the sign of the price reaction during the five minute window [0, +4], we find no evidence of any price reversal in the bad news group during the next 120 minutes. However, while we find a 30 percent price reversal in the good news group by the 30-minute point, there is no further reversal during the next 90 minutes. In addition, we find no inter-day price reversals when we regress *Alert Day* returns against returns of the next two days.[27] Overall, while there is some evidence of intra-day reversal in good news alerts, our primary results do not appear to be driven by noise trading.

To gain further insight into the intra-day market responses to Dow Jones alerts and periodic filings, we examine volume reactions separately for small versus large trades during the window [0, +15]. Following Bhattacharya (2001), we define a large (small) trade as any transaction with a dollar value higher than $50,000 (lower than $5,000). However, when the stock price is over $50, any transaction with less than or equal to 100 shares is also defined as a small trade. We calculate the standardized trading volume (*SVOL*) separately for large and small trades by deflating the corresponding raw volumes by total shares outstanding. Untabulated results show that: (1) both large and small trades exhibit significant volume reactions to the issuance of Dow Jones alerts in the first 16-minute event window; (2) small trade volume reaction in the event window exceeds that in the corresponding pseudo-event window by 83.4 percent, but the reaction is only 35.6 percent higher for large trades; and (3) periodic SEC filings generate insignificant volume reactions from both large and small trades. Overall, the large and small trade results echo the aggregate trading volume results. Finally, we find that large trades neither lead nor lag small trades, further refuting the alternative explanation of noise trading.[28]

Prior research suggests that bad news would attract more coverage by media and other information intermediaries (Gaa 2009), as managers have lower incentives to disseminate such information voluntarily (Kothari et al. 2009). Based on the first five-minute price response to the alert, we assign firm-quarters into bad news, good news, and no response categories. Using an ordered probit model, we find that none of the news categories in Table 6 is significantly associated with the sign of the market response. Taken together, our analysis provides a consistent picture of the market reacting immediately to the release of Dow Jones alerts. While we find no pervasive evidence of an immediate price or volume reaction to periodic reports, our results do not suggest that periodic reports never engender instantaneous market reaction. Future research could

---

[27] In our inter-day analysis, we use a pseudo-event day (i.e., the same weekday in the next week) as the benchmark to control for the short-term price reversal documented in the finance literature (see discussions in Thomas and Zhang [2008]).

[28] While the total trading volume is not correlated with the level of institutional ownership, we find a statistically significant positive (negative) correlation between event-window volume due to large (small) trades and institutional ownership. To the extent that small trades are more representative of individual trading, DJCFA service is acting as an attention-grabber that facilitates trades by individual investors (Barber and Odean 2008).

TABLE 7

Intra-Day Absolute Returns and Trading Volume Surrounding Dow Jones Filing Alerts, SEC Filings, and Control (Pseudo-Event) Periods

Panel A: Dow Jones Filing Alerts

<table>
<tr><th></th><th></th><th colspan="5">Event</th></tr>
<tr><th></th><th></th><th>Event Window [0,+15]</th><th>Pre-Event Window [−16,−1]</th><th>Post-Event Window [+16,+31]</th><th>Event Minus Pre-Event</th><th>Event Minus Post-Event</th></tr>
<tr><th></th><th></th><th>(1)</th><th>(2)</th><th>(3)</th><th>(4)</th><th>(5)</th></tr>
<tr><td>All</td><td>|Return|</td><td>0.485</td><td>0.409</td><td>0.440</td><td>0.075***</td><td>0.044***</td></tr>
<tr><td>(7,615)</td><td>Turnover</td><td>0.029</td><td>0.024</td><td>0.025</td><td>0.005***</td><td>0.004***</td></tr>
<tr><td>Alerts on 10-K</td><td>|Return|</td><td>0.532</td><td>0.397</td><td>0.491</td><td>0.134***</td><td>0.040</td></tr>
<tr><td>(2,704)</td><td>Turnover</td><td>0.027</td><td>0.021</td><td>0.024</td><td>0.006***</td><td>0.003**</td></tr>
<tr><td>Alerts on 10-Q</td><td>|Return|</td><td>0.459</td><td>0.416</td><td>0.412</td><td>0.043**</td><td>0.047***</td></tr>
<tr><td>(4,911)</td><td>Turnover</td><td>0.030</td><td>0.025</td><td>0.026</td><td>0.005***</td><td>0.004***</td></tr>
</table>

<table>
<tr><th></th><th></th><th colspan="5">Pseudo-Event (One Week after Event)</th><th></th></tr>
<tr><th></th><th></th><th>Event Window [0,+15]</th><th>Pre-Event Window [−16,−1]</th><th>Post-Event Window [+16,+31]</th><th>Event Minus Pre-Event</th><th>Event Minus Post-Event</th><th>Event Minus Pseudo-Event</th></tr>
<tr><th></th><th></th><th>(6)</th><th>(7)</th><th>(8)</th><th>(9)</th><th>(10)</th><th>(11)</th></tr>
<tr><td>All</td><td>|Return|</td><td>0.357</td><td>0.380</td><td>0.369</td><td>−0.023*</td><td>−0.012</td><td>0.128***</td></tr>
<tr><td>(7,615)</td><td>Turnover</td><td>0.020</td><td>0.020</td><td>0.019</td><td>0.000</td><td>0.001</td><td>0.009***</td></tr>
<tr><td>Alerts on 10-K</td><td>|Return|</td><td>0.385</td><td>0.402</td><td>0.397</td><td>−0.017</td><td>−0.012</td><td>0.147***</td></tr>
<tr><td>(2,704)</td><td>Turnover</td><td>0.022</td><td>0.021</td><td>0.020</td><td>0.000</td><td>0.001</td><td>0.005**</td></tr>
<tr><td>Alerts on 10-Q</td><td>|Return|</td><td>0.342</td><td>0.368</td><td>0.354</td><td>−0.026*</td><td>−0.012</td><td>0.117***</td></tr>
<tr><td>(4,911)</td><td>Turnover</td><td>0.019</td><td>0.019</td><td>0.018</td><td>−0.001</td><td>0.000</td><td>0.011***</td></tr>
</table>

(continued on next page)

Panel B: SEC Filings Tagged by Alerts

		Event				
		Event Window [0,+15]	**Pre-Event Window [−16,−1]**	**Post-Event Window [+16,+31]**	**Event *Minus* Pre-Event**	**Event *Minus* Post-Event**
		(1)	**(2)**	**(3)**	**(4)**	**(5)**
All	|Return|	0.332	0.333	0.358	0.000	−0.026
(3,669)	Turnover	0.019	0.018	0.020	0.000	−0.002**
10-K Filings	|Return|	0.318	0.337	0.346	−0.019	−0.028
(1,088)	Turnover	0.019	0.018	0.020	0.001	−0.001
10-Q Filings	|Return|	0.339	0.331	0.363	0.008	−0.025
(2,581)	Turnover	0.018	0.018	0.020	0.000	−0.002**

		Pseudo-Event (One Week after Event)					
		Event Window [0,+15]	**Pre-Event Window [−16,−1]**	**Post-Event Window [+16,+31]**	**Event *Minus* Pre-Event**	**Event *Minus* Post-Event**	**Event *Minus* Pseudo-Event**
		(6)	**(7)**	**(8)**	**(9)**	**(10)**	**(11)**
All	|Return|	0.333	0.326	0.353	0.007	−0.020	−0.001
(3,669)	Turnover	0.018	0.018	0.019	0.000	−0.002***	0.001
10-K Filings	|Return|	0.358	0.334	0.378	0.024	−0.020	−0.040
(1,088)	Turnover	0.017	0.016	0.019	0.001	−0.001	0.002
10-Q Filings	|Return|	0.323	0.323	0.343	0.000	−0.020	0.015
(2,581)	Turnover	0.018	0.018	0.020	−0.001	−0.002**	0.001

*, **, *** Represent statistical significance at two-tailed 0.1, 0.05, and 0.01 levels, respectively.

This table is based on the sample period 1997–2004.

Panel A presents the absolute returns and shares turnover during the 16-minute alert window and the corresponding pseudo-event window. All alerts and the two subsamples—10-K-based alerts and 10-Q-based alerts—are shown separately with the respective numbers of observations in parentheses. Only Dow Jones filing alerts issued (1) between 10:30 a.m. and 3:00 p.m. of trading days and (2) beyond 15 minutes after the corresponding SEC filings are included in this analysis. The corresponding pseudo-event window starts with the same hour and minute as the filing alert on the same weekday of the week that immediately follows the filing alert. |Return| is the absolute value of return. Turnover is number of shares traded divided by total shares outstanding. Both metrics have been winsorized at top 0.1 percent and are shown in percentage (%). Event Window refers to the 16 minutes [0, +15] relative to the filing alert. Pre-Event (Post-Event) Window refers to the 16-minutes period immediately before (after) Event Window. The significance tests on difference are based on regressions with the difference as dependent variable regressed on the intercept, after controlling for heteroscedasticity and firm-specific clustering. Panel B presents similar metrics for SEC filings tagged by alerts. Only SEC filings that (1) are filed between 10:30 a.m. and 3:00 p.m. of trading days and (2) are tagged by at least one alert with the first alert issued beyond 15 minutes after the SEC filings are included.

identify circumstances that contribute to immediate processing and reaction to information in periodic reports, which could help regulators grapple with the information mosaic of the mandatory disclosure system currently in place.

VI. CONCLUSION

As capital markets continue the transition into the information age, market participants face not only exponential growth in the quantum of available financial information, but also rapid increases in the number of channels for obtaining information. The new information landscape can stifle even sophisticated investors as they encounter information overload. Consequently, major capital market participants are increasingly relying on internal resources as well as delegated external information intermediaries to screen the influx of corporate financial reports and identify value-relevant information on a timely basis to make informed investment decisions.

Our study focuses on the Corporate Filing Alert service provided by Dow Jones, Inc., which seeks to identify and convey important market-moving information buried in lengthy periodic SEC reports to market participants. We find that for all 10-K/10-Q reports filed between 1997 and 2004, only 9 percent received one or more Dow Jones alerts. Sixty-eight percent of the alerts are issued within 24 hours after the release of corresponding SEC reports.

In general, we posit that Dow Jones' decision to issue an alert is driven by newsworthiness and, therefore, by the perceived value that investors might place on the selected information in periodic SEC reports. We consider investor demand arising from investor awareness and information environment. We find evidence that the likelihood of receiving alerts increases in firms that do not release preliminary earnings, have credit ratings, are included in major market indices, have litigation exposure, or report losses. Among key firm-specific events, firms with a nonstandard audit opinion, in the process of delisting, reporting unusual accounting items, or raising equity capital are more likely to receive an alert. Conditional on issuing an alert, alerts are issued faster on the periodic reports of firm-quarters without a preliminary earnings release or firms from industries with high litigation risk.

Regarding the information content of alerts, we document significant price and volume reactions on the day when news alerts are issued, with bankruptcy- and legal-related news generating the largest reactions. To circumvent any selection bias in firm-quarters followed by DJCFA, we examine the 121-minute event window surrounding the release of alerts, finding that both absolute stock returns and turnover are significantly higher in the 16-minute window immediately following the alert compared with the pre-event or pseudo-event period. The market reactions to alerts are not due to noise trading, although both large and small trades contribute to the reactions. Taken together, our results support the conclusion that Dow Jones alerts convey value-relevant information to investors. Future research could examine whether the documented market reaction reflects both real and informational effects of Dow Jones alerts.

From a regulatory standpoint, the current SEC Commissioner Troy A. Paredes noted that "[t]he federal securities laws primarily focus on … mandating disclosure. Relatively little attention is paid to how the information is used—namely, how investors and securities market professionals search and process information and make decisions based on the information the federal securities laws make available" (Paredes 2003, 418).[29] Our study suggests that the SEC should consider expanding both the breadth and depth of the interactive disclosure requirements. To the extent that mandatory disclosures become more technologically friendly, capital market information intermediaries are bound to leverage the technology by designing faster and better-targeted information search and dissemination strategies to further enhance the efficiency of U.S. capital markets.

---

[29] Commissioner Troy A. Paredes' writing that we cite was published when he was an academic.

REFERENCES

Ajinkya, B., S. Bhojraj, and P. Sengupta. 2005. The association between outside directors, institutional investors and the properties of management earnings forecasts. *Journal of Accounting Research* 43 (3): 343–376.

Ali, A., L. Hwang, and M. A. Trombley. 2003. Arbitrage risk and the book-to-market anomaly. *Journal of Financial Economics* 69 (2): 355–373.

Amir, E., and J. Livnat. 2005. The economic consequences of (not) issuing preliminary earnings announcement. Working paper, London Business School, and New York University.

Balsam, S., E. Bartov, and C. Marquardt. 2002. Accruals management, investor sophistication, and equity valuation: Evidence from 10-Q filings. *Journal of Accounting Research* 40 (4): 987–1012.

Barber, B. M., and T. Odean. 2008. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies* 21 (2): 787–818.

Beyer, A., D. A. Cohen, T. Z. Lys, and B. R. Walther. 2009. The financial reporting environment: Review of the recent literature. Working paper, Stanford University.

Bhattacharya, N. 2001. Investors' trade size and trading responses around earnings announcements: An empirical investigation. *The Accounting Review* 76 (2): 221–244.

Bhattacharya, U., N. Galpin, R. Ray, and X. Yu. 2009. The role of the media in the Internet IPO bubble. *Journal of Financial and Quantitative Analysis* 44 (03): 657–682.

Blom, G. 1958. *Statistical Estimates and Transformed Beta-Variables*. New York, NY: John Wiley and Sons, Inc.

Brown, P., G. Clinch, and G. Foster. 1992. *Market Microstructure and Capital Market Information Content Research, Studies in Accounting Research 32*. Sarasota, FL: American Accounting Association.

Bushee, B., J. E. Core, W. Guay, and S. J. W. Hamm. 2010. The role of the business press as an information intermediary. *Journal of Accounting Research* 48 (1): 1–19.

Bushman, R. M., J. D. Piotroski, and A. J. Smith. 2004. What determines corporate transparency? *Journal of Accounting Research* 42 (2): 207–252.

Carter, M. E., and B. S. Soo. 1999. The relevance of form 8-K reports. *Journal of Accounting Research* 37 (1): 119–132.

Chen, S., M. L. DeFond, and C. W. Park. 2002. Voluntary disclosure of balance sheet information in quarterly earnings announcements. *Journal of Accounting and Economics* 33 (2): 229–251.

Chen, H., G. Noronha, and V. Singal. 2004. The price response to S&P 500 index additions and deletions: Evidence of asymmetry and a new explanation. *The Journal of Finance* 59 (4): 1901–1929.

———, G. Noronha, and V. Singal. 2006. S&P 500 index changes and investor awareness. *Journal of Investment Management* 4 (2): 23–37.

Collins, D. W., E. L. Maydew, and I. S. Weiss. 1997. Changes in the value relevance of earnings and book values over the past forty years. *Journal of Accounting and Economics* 24 (1): 39–67.

———, G. Gong, and P. Hribar. 2003. Investor sophistication and the mispricing of accruals. *Review of Accounting Studies* 8: (2–3): 251–276.

Cready, W. M., and P. G. Mynatt. 1991. The information content of annual reports: A price and trading response analysis. *The Accounting Review* 66 (3): 291–312.

Dawkins, M. C., and L. S. Bamber. 1998. Does the medium matter? The relations among bankruptcy petition filings, broadtape disclosure, and the timing of price reactions. *The Journal of Finance* 53 (3): 1149–1163.

Docking, D. S., and R. J. Dowen. 2006. Evidence on stock price effects associated with changes in the S&P 600 SmallCap Index. *Quarterly Journal of Business and Economics* 45 (1/2): 89–114.

Dopuch, N., R. Holthausen, and R. Leftwich. 1986. Abnormal stock returns associated with media disclosures of 'subject' to qualified audit opinions. *Journal of Accounting and Economics* 8 (2): 93–117.

D'Souza, J., K. Ramesh, and M. Shen. 2010a. Disclosure of GAAP line items in earnings announcements. *Review of Accounting Studies* 15 (1): 179–219.

——— 2010b. The interdependence between institutional ownership and information dissemination by data aggregators. *The Accounting Review* 85 (1): 159–193.

Dyck, A., and L. Zingales. 2003. The media and asset prices. Working paper, Harvard Business School.

Easton, P. D., and M. E. Zmijewski. 1993. SEC form 10K/10Q reports and annual reports to shareholders:

reporting lags and squared market model prediction errors. *Journal of Accounting Research* 31 (1): 113–129.

Ecker, F., J. Francis, P. Olsson, and K. Schipper. 2006. The effect of the amount and configuration of news on inferences about firm-specific events. Working paper, Duke University.

Elliott, W. B., B. F. Ness, M. D. Walker, and R. S. Warr. 2006. What drives the S&P 500 inclusion effect? An analytical survey. *Financial Management* 35 (4): 31–48.

Fang, L., and J. Peress. 2009. Media coverage and the cross-section of stock returns. *The Journal of Finance* 64 (5): 2023–2052.

Foster, G. 1979. Briloff and the capital market. *Journal of Accounting Research* 17 (1): 262–274.

Francis, J., K. Schipper, and L. Vincent. 2002. Expanded disclosures and the increased usefulness of earnings announcements. *The Accounting Review* 77 (3): 515–546.

Frankel, R., M. Johnson, and D. Skinner. 1999. An empirical examination of conference calls as a voluntary disclosure medium. *Journal of Accounting Research* 37 (1): 133–150.

Gaa, C. 2009. Asymmetric attention to good and bad news and the neglected firm effect in stock returns. Working paper, University of Oregon.

Gerety, M. S., and H. J. Mulherin. 1992. Trading halts and market activity: An analysis of volume at the open and the close. *The Journal of Finance* 47 (5): 1765–1784.

Gilson, S. C. 1995. Investing in distressed situations: A market survey. *Financial Analysts Journal* 51 (6): 8–27.

Griffin, P. A. 2003. Got information? Investor response to form 10-K and form 10-Q EDGAR filings. *Review of Accounting Studies* 8 (4): 433–460.

Hayn, C. 1995. The information content of losses. *Journal of Accounting and Economics* 20 (2): 125–153.

Hirshleifer, D. A., J. N. Myers, L. A. Myers, and S. H. Teoh. 2008. Do individual investors cause post-earnings announcement drift? Direct evidence from personal trades. *The Accounting Review* 83 (6): 1521–1550.

Hong, H., and J. C. Stein. 1999. A unified theory of underreaction, momentum trading, and over-reaction in asset markets. *The Journal of Finance* 54 (6): 2143–2184.

In, F., S. Sadique, and M. Veeraraghavan. 2007. Spin in earnings information and stock returns. Working paper, Monash University.

Jennings, R. H., and C. B. Barry. 1983. Information dissemination and portfolio choice. *Journal of Financial and Quantitative Analysis* 18 (1): 1–19.

Jiambalvo, J., S. Rajgopal, and M. Venkatachalam. 2002. Institutional ownership and the extent to which stock prices reflect future earnings. *Contemporary Accounting Research* 19 (1): 117–145.

Kim, S. T., J.-C. Lin, and M. B. Slovin. 1997. Market structure, informed trading, and analysts' recommendations. *Journal of Financial and Quantitative Analysis* 32 (4): 507–524.

Kothari, S. P., S. Shu, and P. D. Wysocki. 2009. Do managers withhold bad news? *Journal of Accounting Research* 47 (1): 241–276.

Li, E. X., and K. Ramesh. 2009a. Market reaction surrounding the filing of periodic SEC reports. *The Accounting Review* 84 (4): 1171–1208.

———, and ———. 2009b. Needles shining in a haystack: What draws investors' attention to corrective information in periodic SEC filings. Working paper, University of Rochester, and Michigan State University.

Lu, H. 2006. Information diffusion and the boundary of market efficiency: Empirical evidence from earnings announcements. Working paper, University of Toronto.

Mase, B. 2006. Investor awareness and the long-term impact of FTSE 100 index redefinitions. *Applied Financial Economics* 16 (15): 1113–1118.

Mashruwala, C. A., S. Rajgopal, and T. J. Shevlin. 2006. Why is the accrual anomaly not arbitraged away? The role of idiosyncratic risk and transaction costs. *Journal of Accounting and Economics* 42 (1–2): 3–33.

Merton, R. C. 1987. A simple model of capital market equilibrium with incomplete information. *The Journal of Finance* 42 (3): 483–510.

Miller, G. S. 2006. The press as a watchdog for accounting fraud. *Journal of Accounting Research* 44 (5): 1001–1033.

Ng, J., T. Rusticus, and R. Verdi. 2008. Implications of transaction costs for the post-earnings-announcement drift. *Journal of Accounting Research* 46 (3): 661–696.

O'Brien, P. C., and R. Bhushan. 1990. Analyst following and institutional ownership. *Journal of Accounting Research* 28 (Supplement): 55–76.

Paredes, T. A. 2003. Blinded by the light: Information overload and its consequences for securities regulation. *Washington University Law Quarterly* 81: 417–485.

Piotroski, J. D., and D. Roulstone. 2004. The influence of analysts, institutional investors, and insiders on the incorporation of market, industry and firm-specific information into stock prices. *The Accounting Review* 79 (4): 1119–1151.

Romer, P. 1990. Endogenous technological change. *The Journal of Political Economy* 98 (S5): 71–102.

Rose, F. 1999. *The Economics, Concept, and Design of Information Intermediaries.* Heidelberg, NY: Physica-Verlag.

Solomon, D. 2009. Selective publicity and stock prices. Working paper. University of Southern California.

Soltes, E. 2009. News dissemination and the impact of the business press. Working paper, The University of Chicago.

Thomas, J., and F. Zhang. 2008. Over-reaction to intra-industry information transfers. *Journal of Accounting Research* 46 (4): 909–940.

Veldkamp, L. 2006. Media frenzies in markets for financial information. *The American Economic Review* 96 (3): 577–601.

Wooldridge, J. 2002. *Econometric Analysis of Cross Section and Panel Data.* Cambridge, MA: MIT Press.

——— 2003. *Introductory Econometrics: A Modern Approach.* 2nd edition. Cincinnati, OH: South-Western College Publishing.

Wurgler, J., and E. Zhuravskaya. 2002. Does arbitrage flatten demand curves for stocks? *The Journal of Business* 75 (4): 583–608.

You, H., and X. Zhang. 2009. Financial reporting complexity and investor underreaction to 10-K information. *Review of Accounting Studies* 14 (4): 559–586.

THE ACCOUNTING REVIEW
Vol. 86, No. 2
2011
pp. 703–733

American Accounting Association
DOI: 10.2308/accr.00000024

Stock Repurchases and Executive Compensation Contract Design: The Role of Earnings per Share Performance Conditions

Steven Young
Lancaster University Management School

Jing Yang
Towers Perrin

ABSTRACT: We examine the link between firms' stock repurchase activity and the presence of earnings per share (EPS) performance conditions in executive compensation contracts. Findings reveal a strong positive association between repurchases and EPS-contingent compensation arrangements. Further analysis suggests net benefits to shareholders from this association. Specifically, repurchasers experience larger increases in total payouts; the positive association between repurchases and cash performance is more pronounced for firms with EPS targets in the presence of surplus cash; undervalued firms with EPS targets are more likely to signal mispricing through a repurchase; and repurchasers with EPS conditions are associated with lower abnormal accruals. We find no evidence that EPS-driven repurchases impose costs on shareholders in the form of investment myopia.

Keywords: *corporate payout policy; performance targets; earnings management; efficient contracting.*

Data Availability: *The data are available from public sources identified in the study.*

We are indebted to Paulo Alves for excellent research support. We are also grateful for the insightful suggestions and support offered by two anonymous reviewers. Helpful comments and suggestions were also provided by Amir Amel-Zadeh, Mary Ellen Carter, John Forker, Stephen X. H. Gong, Marwan Izzeldin, Mike Jones, Steven Kachelmeier (editor), and seminar participants at the University of Bath, University of Bristol, ESSEC, University of Mannheim, Stockholm School of Economics, the 2009 Annual Meeting of the American Accounting Association, the 2009 European Financial Management Symposium on Corporate Governance and Control, the 2007 Annual Meeting of the British Accounting Association, and the 2006 and 2009 Financial Reporting and Business Communication Conferences. Financial support was provided by The Leverhulme Trust (Project F/00185/I).

All opinions expressed herein are attributable solely to the authors and in no way reflect the views of Towers Perrin.

Editor's note: Accepted by Steven Kachelmeier.

Submitted: October 2008
Accepted: August 2010
Published Online: March 2011

I. INTRODUCTION

This study investigates the link between firms' stock repurchase activity and the presence of earnings per share performance conditions in executive compensation contracts. Our analysis seeks to address the apparent disconnect between theory and practice regarding repurchases. One the one hand, traditional academic theories identify factors such as signaling (Vermaelen 1981), agency costs (Fenn and Liang 2001), and leverage (Dittmar 2000) as important determinants of repurchase activity. On the other hand, survey and anecdotal evidence highlight earnings per share (EPS) as a key factor influencing managers' repurchase decisions (Badrinath and Varaiya 2000; Brav et al. 2005; Caster et al. 2006). Exploring why managers attach such weight to the EPS impact of repurchases represents an important step toward a better understanding of this increasingly significant aspect of payout policy.

Recent research sheds light on the links between repurchase decisions and EPS-related considerations. Kahle (2002), Bens et al. (2002), and Bens et al. (2003) focus on the dilutive impact of employee stock options (ESOs). Their findings suggest repurchases are a managerial response to EPS dilution concerns. Evidence also suggests that managers use repurchases for benchmark-beating purposes, including meeting or exceeding analysts' EPS forecasts (Hribar et al. 2006), preserving a sequence of EPS improvement (Myers et al. 2007), and maintaining historic EPS growth rates (Bens et al. 2003).

Our analysis builds on prior research by examining whether managers' stock repurchase decisions are sensitive to explicit EPS-related incentives provided by executive compensation contracts. Compensation contracts linking rewards to EPS performance provide executives with direct and potentially powerful incentives to manage reported EPS. We therefore test whether repurchase activity is higher for firms with executive compensation tied to EPS performance.

Empirical tests employ data for a comprehensive sample of U.K. nonfinancial firms over the period January 1998 through December 2006. Several features make the U.K. a particularly attractive setting in which to explore the link between repurchases and compensation contract design. First, in addition to executive bonus plans that routinely condition rewards on EPS performance, executives' long-term incentives including options and restricted stock frequently employ EPS vesting conditions (Conyon et al. 2000; Carter et al. 2009). Second, regulatory restrictions governing the treatment of repurchases during the majority of our sample period help simplify empirical tests by tempering the link between repurchase activity and the dilutive effects of ESOs. In particular, U.K. Company Law required repurchased shares to be cancelled immediately rendering repurchases a costly device for offsetting ESO-related EPS dilution (because issuing new shares is administratively more costly than reissuing treasury stock). Instead, U.K. firms with ESO programs typically established a wholly owned trust company to repurchase (and reissue) shares on their behalf. Under U.K. GAAP, shares held by ESO trusts are excluded from the EPS calculation until shares vest unconditionally. Since ESO shares are purchased solely to fund share-based compensation plans and because ESO purchases do not meet the legal definition of a stock repurchase, our tests can distinguish between repurchases driven by dilution concerns (ESO shares) and repurchases driven by other factors.[1]

Findings reveal a significant association between repurchase activity and the presence of EPS-based compensation arrangements. The predicted odds of a repurchase for firms for which

[1] U.K. Company Law was amended in December 2003 to allow firms to hold repurchased shares as treasury stock. Consistent with the absence of ESO-related motives for repurchases, none of our sample firms mentioned the dilutive impact of stock-based compensation plans among the list of repurchase reasons disclosed in their annual reports prior to this date. Following the regime switch, most U.K. firms continue to cancel repurchased shares. As described in Section III, we exclude from our subsequent empirical tests firms that explicitly repurchase stock into treasury post-December 2003 to fund option plans.

executive compensation depends on EPS performance are almost twice the level observed for firms for which rewards are independent of EPS. EPS conditions in bonus and option plans are associated with incrementally significant effects. Our findings suggest that EPS performance conditions represent an important determinant of U.K. managers' stock repurchase decisions.

An important ancillary question raised by our findings is how the link between repurchases and executive compensation arrangements impacts shareholder value. We explore this issue by examining the costs and benefits associated with compensation-driven repurchases. On balance, our evidence suggests that EPS-driven repurchases yield net benefits to shareholders. First, repurchasers make higher aggregate payouts to shareholders regardless of the performance conditions employed in executive compensation contracts, and we find no evidence that EPS-driven repurchase payouts occur at the expense of either investment myopia (Bens et al. 2002) or dividend substitution (Grullon and Michaely 2002). Second, tests reveal several contracting benefits from repurchase incentives caused by linking compensation to EPS performance, including a stronger association between cash performance and repurchases in the presence of surplus cash, a higher propensity to signal undervaluation through a repurchase when stock price falls below intrinsic value, and lower accrual manipulation.

Our study contributes to prior research in several ways. First, while extant work links repurchases to the dilutive impact of employee option plans (Kahle 2002; Bens et al. 2002; Bens et al. 2003), and EPS-based bonus plans to the dilutive impact of new equity issues (Huang et al. 2010), the association between repurchase activity and EPS-contingent compensation has attracted little attention. Our study highlights EPS-based executive compensation contracts as an important determinant of repurchase activity that is entirely consistent with the EPS benefits of buybacks frequently highlighted by management (Badrinath and Varaiya 2000; Brav et al. 2005; Caster et al. 2006). Our findings complement and extend Marquardt et al.'s (2009) evidence that EPS-based bonus plans explain managers' preference for accelerated stock repurchases over regular open market buybacks. We add to their findings by demonstrating that EPS-based compensation arrangements also explain the underlying decision to repurchase stock and that shareholders benefit from this relation. Second, our analysis relates to work on performance measure choice in compensation contracts. In particular, prior research emphasizes how EPS targets encourage overinvestment (Brealey et al. 2008, 889). Our results provide a counterbalance to this view by highlighting how the repurchase incentives created by EPS-contingent pay help align managers' interests with those of shareholders. Our conclusion is consistent with Huang et al. (2010) who model the use of EPS conditions in executive bonus plans and find that EPS-contingent compensation helps resolve agency problems by protecting current shareholders from a reduction in proportional ownership. Our analysis also contributes to the small body of research exploring the interaction between alternative earnings management instruments (e.g., Demski et al. 2004; Cohen et al. 2008). We argue that shareholders gain when managers manipulate EPS through repurchases rather than accounting accruals, and we show that the incentive created by EPS targets to manipulate via repurchases correlates with less accrual management.

Section II develops the link between repurchases and executive compensation arrangements, and reviews the structure of executive compensation plans in the U.K. Section III provides details of our sample, data, and research design. Section IV reports results of tests that examine the association between repurchases and EPS-based performance targets in executive compensation plans. Section V explores the implications for shareholders of repurchases motivated by EPS-contingent performance conditions. Section VI concludes.

II. LITERATURE REVIEW AND INSTITUTIONAL OVERVIEW

Stock Repurchases, Earnings Management, and Executive Compensation

Despite being overlooked in much of the corporate payout policy literature, managers and financial commentators have long recognized the EPS implications of stock repurchases. The impact of repurchases on reported EPS represents the net of both numerator and denominator effects. The numerator effect, which works to reduce EPS, represents the decline in earnings caused by an increase in borrowing (for repurchases financed with debt) or a reduction in investment returns (for repurchases financed using cash reserves). The denominator effect serves to increase EPS by reducing the number of shares outstanding. Repurchases have a positive net effect on EPS for firms whose earnings-to-price ratio exceeds the opportunity cost of funds (i.e., either the after-tax return on short-term cash investments or the cost of debt; Guay 2002; Bens et al. 2003; Hribar et al. 2006). Survey evidence reported by Brav et al. (2005) highlights the central role EPS considerations play in shaping managers' repurchase decisions, with three-quarters of senior executives citing EPS growth as an important factor affecting their repurchase decisions.

Research has begun to explore the link between repurchases and EPS in several contexts. One EPS-related factor predicted to motivate repurchases is earnings dilution caused by ESO plans. Accretive stock repurchases can offset the dilutive effects of ESOs on reported EPS in several ways. For example, while ESO exercises reduce basic EPS by increasing the weighted average number of shares outstanding for the period, managers can mitigate this dilution by repurchasing shares to fund option exercises. Bens et al. (2002) and Kahle (2002) present evidence consistent with this option-funding hypothesis. Conversely, Bens et al. (2003) conclude that repurchases are not a response to the dilutive impact of option exercises on basic EPS. Instead, their results suggest the link between repurchases and options is driven by the effect of ESOs on diluted EPS, with buybacks increasing in the level of in-the-money ESOs outstanding.

Repurchases motivated by EPS considerations have also been linked with benchmark-beating earnings management activity. Controlling for dilution effects, Bens et al. (2003) find that repurchases are increasing in the amount by which earnings undershoot the level required to sustain historical diluted EPS growth. Myers et al. (2007) document similar behavior in a sample of firms characterized by long strings of consecutive quarterly EPS increases, with managers strategically timing repurchases to boost reported EPS when the string would otherwise be broken. Meanwhile, Hribar et al. (2006) conclude that managers use repurchases to meet or beat analysts' consensus EPS forecasts. In all these studies, benchmark-beating stock repurchase activity is motivated through implicit managerial incentives in the form of higher stock-related compensation, greater job security, and a lower cost of capital.[2]

Compensation contracts represent a powerful source of incentives for managers. For example, research demonstrates that executives use their accounting discretion to manipulate earnings in response to compensation-driven considerations (Bushman and Smith 2001). Further, a growing body of evidence suggests that corporate payout decisions are sensitive to executives' compensation arrangements. For instance, firms for which the executives' annual bonus pool is contingent on dividends paid are associated with higher dividend payouts and yields (White 1996), while ESOs that are not dividend-protected create incentives for executives to reduce dividend payments

---

[2] Bens et al. (2003, 75–76) argue that explicit compensation contract considerations are not the source of their findings linking repurchases to EPS manipulation. They estimate firm-specific correlations between CEO cash compensation and reported EPS and use the median correlation to divide firms into high and low cash compensation-EPS sensitivity firms. Tests reveal no evidence that their main results differ across the two subsamples. However, this approach is unlikely to provide a powerful means of distinguishing firms with explicit EPS targets from those using alternative earnings-based performance metrics. Further, since the approach focuses exclusively on cash compensation, it ignores the impact of EPS vesting conditions in long-term compensation.

(Lambert et al. 1989; Kahle 2002; Fenn and Liang 2001). In related research, Aboody and Kasznik (2008) demonstrate how compensation plan design can help align managers' cash payout decisions with shareholders' tax-driven payout preferences. Finally, Wallace (1997) and Marquardt et al. (2009) examine the link between bonus plan performance conditions and corporate payout policy. Using a sample of firms adopting residual income-based plans (which penalize managers for accumulating capital that earns less than the opportunity cost of capital), Wallace (1997) documents a post-adoption rise in repurchase activity as managers liquidate unproductive assets. Marquardt et al. (2009), meanwhile, provide direct evidence on the link between EPS performance conditions and stock repurchase activity. Specifically, they find that managers are more likely to favor accelerated stock repurchases (which record the full EPS benefits of the repurchase immediately) over regular open market repurchases when their bonus plans are explicitly tied to EPS performance. However, the extent to which EPS-based compensation arrangements explain the propensity to repurchase more generally remains unexplored.

EPS is a popular performance metric used in executive compensation contracts (Murphy 1999; Conyon et al. 2000; Pass et al. 2000). Compensation contracts that tie managerial rewards to EPS create explicit incentives for executives to manage the EPS denominator through repurchases (over and above any implicit market-based incentives associated with increasing stock-based wealth and improving job security). These direct incentives are absent in compensation contracts that employ non-per-share-based earnings metrics such as return on assets, and nonaccounting measures such as stock price or qualitative targets linked to personal objectives. Accordingly, we predict that stock repurchase activity will be positively associated with the incidence of EPS-based performance conditions in executive compensation contracts. We test this prediction in the U.K. where short- and long-term elements of executive pay are linked to EPS (Conyon et al. 2000; Pass et al. 2000; Carter et al. 2009).

Overview of Executive Compensation Arrangements in the U.K.

The typical compensation package for a U.K. executive director includes both short-term bonus arrangements and longer-term incentives such as options and restricted stock (Conyon and Murphy 2000). Bonus payments are normally linked to short-term performance measures and objectives. In addition, U.K. firms regularly impose performance-vesting conditions in executives' long-term stock- and cash-based plans. Widespread adoption of performance-vesting conditions in executives' long-term compensation plans can be traced to the Greenbury Report (1995), which proposed that all long-term incentives (including option plans) should include challenging performance criteria. After December 31, 1995, revised London Stock Exchange rules required all listed firms to either comply with the Greenbury recommendation or publish a statement explaining noncompliance. Further pressure to adopt performance conditions for long-term incentives was applied by influential shareholder groups including the Association of British Insurers and the National Association of Pension Funds. As a result, performance vesting conditions in long-term plans are now commonplace among U.K. firms (Carter et al. 2009).

While best practice compensation guidelines do not favor any single performance metric, survey evidence reveals widespread adoption of EPS-based targets. For example, Conyon et al. (2000) report that 72 percent of option plans with performance-contingent vesting conditions define targets in terms of EPS growth, while Pass et al. (2000) find that 34 percent of long-term incentive plans (LTIPs) surveyed had an EPS performance condition. Accordingly, the performance conditions applied in long-term compensation arrangements often mirror those used in short-term bonus plans, for which EPS targets have long been used. The widespread use of EPS targets in executive compensation arrangements is expected to create powerful incentives for U.K. executives to manage EPS realizations through repurchases.

III. DATA AND METHODS

Sample and Data

The initial sampling frame comprises all U.K.-resident firms (excluding closed-end investment trusts) listed on the London Stock Exchange (LSE) with fiscal year-ends between January 1, 1998 and December 31, 2006. The sample period starts in 1998 because executive compensation data are collected with a one-year lag and disclosures relating to performance conditions in executive compensation contracts are patchy before 1997.

Firm-level repurchase data relate to aggregate reacquisitions made during a fiscal year. Only repurchases executed in the open market or via self-tender offer are used in subsequent tests. We hand-collect annual repurchase data from firms' published financial statements. This process involves identifying potential repurchasers using a variety of news sources including the London Stock Exchange Regulatory News Service, the Securities Data Corporation, and *The Financial Times*. Financial statements with year-ends between January 1998 and December 2006 are then examined for all firms in the provisional list to identify the number, value, and fraction of shares repurchased. The resulting sample comprises 1,047 repurchase firm-year observations for 460 firms. We remove financial firms due to the unique nature of performance measurement in that sector. Utility firms are also removed due to a lack of nonrepurchasing firms in the same sector for matching purposes (see below). A further 67 observations are lost due to missing data required to construct one or more of our test variables. We also exclude ten treasury stock repurchase observations driven entirely by outstanding option commitments. The final sample consists of 665 repurchase firm-years.

Repurchasing firms are drawn from 31 Datastream level-4 nonfinancial industry groups, with no single industry accounting for more than 12 percent of the final sample. The aggregate value of shares reacquired during the sample window exceeds £83 billion, with an average (median) annual repurchase value of £124.9 million (£3.9 million). Repurchase activity in the U.K. is increasing over time, with the aggregate amount rising from £636 million in 1998 to almost £28 billion in 2006. The average (median) annual repurchase involves approximately 5 (3) percent of common shares outstanding. Repurchased shares are cancelled in the majority of cases: only 66 observations (10 percent) utilize the treasury stock option.

Empirical tests require details of performance conditions used in executive compensation contracts, data for which are also hand-collected from firms' published annual reports and financial statements. Collecting such data for all LSE-listed nonrepurchase firms is infeasible. We therefore employ a case-control matched sample design whereby each of the 665 repurchase firm-year observations is paired with a time-, industry- and size-matched nonrepurchasing firm.[3] Matching by industry (Datastream level-4) helps control for factors that are expected to affect payout policy (Smith and Watts 1992) and compensation arrangements (Antle and Smith 1986), while matching by size (lagged total assets) helps to control for established associations between firm size and repurchase activity (Dittmar 2000; Jagannathan et al. 2000), and between firm size and compensation arrangements (Pass et al. 2000). Nonrepurchase control firms are matched with repurchasers at the fiscal year-end immediately preceding the repurchase year. Nonrepurchasers must not have implemented a buyback at any point prior to the matching year or during the subsequent four-year period.

---

[3] Case-control matching unavoidably leads to disproportionate random sampling on the dependent variable. However, subsequent tests linking repurchase activity to EPS-based performance conditions in executive compensation contacts employ logistic regression, a well-known property of which is that slope coefficients remain unbiased in the presence of disproportionate random sampling on the dependent variable (Prentice and Pyke 1979).

Details of the following performance-related elements of executive compensation are collected from repurchase and nonrepurchase firms' annual reports in the matching year: short-term bonus plans, option plans, and long-term incentive plans. Bonus plans comprise all arrangements for which rewards are tied to short-term (≤one-year) performance targets. Option plans comprise all stock-based arrangements granting executives the right to acquire shares at a nonzero exercise price. (Firm-wide employee option plans and save-as-you-earn schemes are excluded.) LTIPs consist of all remaining long-term compensation arrangements not classified as options (e.g., deferred bonus schemes, share matching schemes, zero strike price options, stock appreciation rights, long-term cash-based bonus plans, etc.). We record the performance conditions for all active plans in each category. The data-collection process has to confront two disclosure problems. First, a handful of firms fail to unambiguously disclose use of one or more of the three plan types. Second, some firms fail to provide details of the performance conditions used in one or more of their plans. We use previous years' Annual General Meeting resolutions and remuneration disclosures up to two years ahead to verify plan existence and determine performance conditions employed.[4] Cases for which we cannot unambiguously determine plan existence or the use of an EPS performance condition are coded as nondisclosers.

Research Design

We expect EPS performance conditions to be more prevalent among repurchasing firms. We test this prediction using a conditional logistic regression to model the probability of a repurchase and a left-censored tobit regression to model the value of repurchases:

$$Log\left(\frac{p_{it}}{1-p_{it}}\right) = \gamma_1 NDISC_{ijt-1} + \gamma_2 EPS_{ijt-1} + \sum_{k=1}^{K} \theta_k Controls_{kit-1}, \tag{1}$$

$$Rep_{it} = \lambda_0 + \lambda_1 NDISC_{ijt-1} + \lambda_2 EPS_{ijt-1} + \sum_{k=1}^{K} \delta_k Controls_{kit-1} + \upsilon_{it}. \tag{2}$$

For the conditional logistic model presented in Equation (1), p_{it} is the latent probability that firm i repurchases shares in year t ($y_{it} = 1$). The conditional logistic model is the appropriate estimation method for the matched pair structure of our data (Allison 1999, 203).[5] For the left-censored tobit model presented in Equation (2), *Rep* is the observed value of the latent propensity to repurchase stock (Rep^*): $Rep_{it} = 0$ if $\mathrm{Rep}_{it}^* \leq 0$ and Rep_{it} = value of stock repurchases in fiscal year t scaled by lagged total assets if $Rep_{it}^* > 0$.

The vector of explanatory variables is the same in Equations (1) and (2): *NDISC* is an indicator variable taking the value of 1 if the presence of an EPS performance condition is indeterminate for at least one of the j compensation components (j = bonus plans, stock option plans, and LTIPs), and 0 otherwise; *EPS* is an indicator variable taking the value of 1 if an EPS performance condition is used in at least one of the j compensation components, and 0 otherwise; and *Controls* is a vector of K additional factors expected to influence the repurchase decision.[6] The

[4] Compensation disclosures in the U.K. improved dramatically during the late 1990s and early 2000s. When using one- and two-period-ahead remuneration disclosures, we are careful to distinguish between established plans and new plans introduced subsequently.

[5] We also estimated Equation (1) using a pooled (unmatched) logistic model with very similar results. Results are available from the authors on request.

[6] Incomplete disclosure of performance conditions means that EPS realizations may take one of three forms: EPS condition is used and disclosed; EPS condition is unambiguously not used; and EPS condition is indeterminate due to insufficient disclosure. Defining *EPS* in Equations (1) and (2) as a three-way categorical variable imposes a linearity constraint on the data. The alternative (unconstrained) approach is to recode the three-way variable as three separate

set of control variables includes the market-to-book ratio, net leverage, dividend yield, prior-period abnormal stock price performance, and firm size (Stephens and Weisbach 1998; Dittmar 2000; Grullon and Michaely 2002). Since repurchases have a positive net effect on EPS when a firm's earnings-to-price (E/P) ratio exceeds its opportunity cost of funds, only firms that meet such a condition will repurchase shares to boost EPS. We therefore include an indicator variable for firms reporting negative earnings on the grounds that the E/P condition is least likely to hold in such cases. Prior research documents a link between ESO plans and stock repurchases by U.S. firms (Fenn and Liang 2001; Dittmar 2000). Although U.K. regulatory rules governing stock repurchases militate against such behavior during our sample period, for completeness we include the total number of options outstanding for all employees (scaled by market capitalization) as an additional control variable. To ensure that our analysis is capturing effects unique to EPS, we also control for the presence of earnings-based performance conditions of any description.[7] Finally, we control for the well-established link between repurchases and surplus cash (Stephens and Weisbach 1998; Dittmar 2000; Guay and Harford 2000; Jagannathan et al. 2000). The presence of surplus cash is captured using both stock (surplus cash holdings) and flow (excess cash flow) measures. Our measure of surplus cash holdings is cash and cash equivalents in excess of the level required for normal operations and investments. Following Opler et al. (1999), we estimate surplus cash holdings as the residual from an OLS regression of cash holdings (scaled by lagged noncash assets) on a vector of explanatory variables comprising the market-to-book ratio, net working capital (scaled by lagged total assets), lagged operating cash flow (scaled by lagged total assets), net leverage, R&D spending (scaled by total revenue), the natural logarithm of market capitalization, an indicator variable for nonzero dividend payments, and industry fixed effects. The regression is estimated annually using all Extel nonfinancial firms with complete data after excluding the top and bottom percentiles of scaled cash holdings. We extract the residuals (ε_{it}) from the annual estimations to construct an indicator variable equal to 1 when $\varepsilon_{it} > 0$, and 0 otherwise.

We use two measures of excess cash flow, one based on operating activities (*Free cash flow*) and one based on nonoperating activities (*Excess investing cash flow*). Following Opler and Titman (1993) and Fenn and Liang (2001), *Free cash flow* is an indicator variable equal to 1 for firms with a market-to-book ratio less than the Extel annual sample median and operating cash flow (scaled by lagged total assets) greater than the Extel annual sample median, and 0 otherwise. *Excess investing cash flow* is an indicator variable taking the value of 1 when net cash inflow from investing activities is positive, and 0 otherwise. (Investing cash inflows result from the sale of fixed and intangible assets, associates and other investments, and subsidiaries.) All explanatory variables in Equations (1) and (2) are measured at the start of the repurchase year.

IV. RESULTS

Descriptive Statistics

Table 1, Panel A reports descriptive statistics for the incidence of bonus plans, option plans, and LTIPs. Frequency counts reported in columns 2–4 reveal most firms operate at least one bonus plan and one option plan, whereas only 40 percent of firms have an active LTIP. Cross-sample comparisons indicate that repurchasers are marginally more likely to operate a bonus plan ($p = 0.09$). Consistent with the absence of powerful ESO-related motives for stock repurchases in

---

dummy variables and then use two of these in place of the original variable (Allison 1999, 128–130). Tests reveal that imposing the linearity constraint on our data leads to a reduction in model fit (the change in the likelihood ratio statistic is significant at the 0.01 level), suggesting that the unconstrained formulation presented in Equations (1) and (2) is more appropriate in our case.

[7] For example, residual income also creates incentives to distribute capital earning less than the opportunity cost of funds (Wallace 1997).

TABLE 1

Summary Statistics and Features of Compensation Plan Components for Repurchaser and Nonrepurchaser Matched Pairs

Panel A: Frequency of Plans

	Plan Status by Firm			**Number of Plans**				
Compensation Component	**≥1 Plan**	**No Plan**	**Not Disclosed**	**n**	**Mean**	**Std. Dev.**	**Median**	**Max**
Bonus Plans								
Repurchasers	601	49	15	605	0.931	0.277	1	2
Nonrepurchasers	596	68	1	600	0.904	0.315	1	2
p-value for difference	0.09				0.17		0.17	
Option Plans								
Repurchasers	591	69	5	759	1.150	0.621	1	3
Nonrepurchasers	601	64	0	789	1.186	0.620	1	3
p-value for difference	0.62				0.29		0.31	
Long-Term Incentive Plans								
Repurchasers	282	377	6	377	0.572	0.745	0	3
Nonrepurchasers	287	378	0	373	0.561	0.744	0	3
p-value for difference	0.89				0.77		0.74	

Panel B: Performance Measures by Plan

	Frequency Counts			
	Repurchasers		**Nonrepurchasers**	
Performance Measures by Compensation Component	**n**	**%**	**n**	**%**
Bonus Plans				
Earnings per share	168	(27.8)	121	(20.2)
Profit before tax/EBIT/Operating profit	303	(50.1)	329	(54.8)
Residual income	15	(2.5)	7	(1.2)
Return on capital	21	(3.5)	34	(5.7)
Share price/Total shareholder return	16	(2.6)	16	(2.7)
Personal objectives	88	(14.5)	118	(19.7)
Other	198	(32.7)	252	(42.0)
Not disclosed	153	(25.3)	122	(20.3)
Total Number of Plans	605		600	
Option Plans				
Earnings per share	488	(64.3)	398	(50.4)
Profit before tax/EBIT/Operating profit	24	(3.2)	15	(1.9)
Return on capital	4	(0.5)	1	(0.1)
Share price/Total shareholder return	62	(8.2)	125	(15.8)
Personal objectives	0	(0.0)	1	(0.1)
Other	8	(1.1)	12	(1.5)
No performance condition	176	(23.2)	238	(30.2)
Not disclosed	27	(3.6)	9	(1.1)
Total Number of Plans	759		789	
Long-Term Incentive Plans				
Earnings per share	155	(41.1)	151	(40.5)
Profit before tax/EBIT/Operating profit	13	(3.4)	17	(4.6)

(continued on next page)

Panel B: Performance Measures by Plan

Performance Measures by Compensation Component	Frequency Counts: Repurchasers n	Repurchasers %	Nonrepurchasers n	Nonrepurchasers %
Residual income	6	(1.6)	2	(0.5)
Return on capital	10	(2.7)	13	(3.5)
Share price/Total shareholder return	217	(57.6)	196	(52.5)
Other	14	(3.7)	17	(4.6)
No performance condition	52	(13.8)	43	(11.5)
Not disclosed	1	(0.3)	3	(0.8)
Total Number of Plans	377		373	

The sample comprises 665 fiscal years between January 1, 1998 and December 31, 2006 in which firms repurchased shares and 665 nonrepurchase firm-years matched by fiscal year, industry, and lagged total assets.

All compensation data relate to compensation contracts for executive directors. Data are collected on all plans for the following three elements in executives' compensation contracts: bonuses, share options, and long-term incentives. Bonus plans comprise all arrangements where rewards are tied to short-term (≤one-year) performance targets. Option plans consist of incentive contracts granting executives the right to acquire their firm's shares a non-zero exercise price. (Firm-wide employee share option plans and save-as-you-earn schemes are not included.) Long-term incentive plans (LTIPs) consist of all remaining long-term compensation arrangements not classified as options. Probability values reported in Panel A are for Chi-square tests (column 2), paired t-tests (column 6), and paired Wilcoxon tests (column 8).

For each compensation component, the sum of performance measure percentages reported in Panel B may exceed 100 because some firms use multiple measures in a single plan.

the U.K., the frequency of firms with at least one option plan is equivalent in the two samples, as is the incidence of LTIPs. The final three columns in Panel A report summary statistics for the number of active plans. No significant differences are apparent between the two samples.

Plan-level details of performance conditions are reported in Table 1, Panel B. For bonus plans, 28 percent of the 605 plans operated by repurchasers have EPS performance conditions, as compared to only 20 percent of 600 comparable plans operated by nonrepurchasers. In contrast, aggregate profit-based metrics such as profit before tax, personal objectives, and other measures (e.g., operations metrics, KPIs) are more common among nonrepurchasers. Note also that repurchase firms are characterized by poorer disclosure of bonus-related performance conditions: 25 percent of plans in the repurchase sample contain no details of performance conditions, as compared to only 20 percent of plans in the nonrepurchase sample. Similar patterns are apparent for option plans. Option exercise is conditional on EPS performance in 64 percent of repurchase firms' plans, as compared to 50 percent of plans in the nonrepurchase sample. Results for bonus and option plans provide preliminary evidence that repurchase activity is increasing in the presence of EPS performance conditions. No difference in the use of EPS conditions exists for LTIPs; roughly 40 percent of plans in both samples are conditional on EPS performance.

Summary statistics for our main test variables are reported in Table 2. Seventy-one percent of repurchasers have at least one plan linking at least one element of executive compensation to EPS. The comparable figure for nonrepurchasers is 60 percent, which is significantly lower based on a paired Wilcoxon test ($p < 0.01$). Analyzing the incidence of EPS targets for each compensation element separately reveals that repurchasers are significantly more likely to have at least one bonus plan and at least one option plan tied to EPS. In contrast, repurchase and nonrepurchase firms are equally likely to have an LTIP conditional on EPS. Significant differences across repur-

TABLE 2

Descriptive Statistics

	Repurchase Sample (n = 665)						
	Mean	**Std. Dev.**	**Max**	**Q3**	**Median**	**Q1**	**Min**
Compensation Variables							
EPS	0.71	0.46	1	1	1	0	0
NDISC	0.14	0.35	1	0	0	0	0
EPS_{Bonus}	0.25	0.43	1	1	0	0	0
$NDISC_{Bonus}$	0.24	0.43	1	0	0	0	0
EPS_{Option}	0.62	0.49	1	1	1	0	0
$NDISC_{Option}$	0.05	0.21	1	0	0	0	0
EPS_{LTIP}	0.20	0.40	1	0	0	0	0
$NDISC_{LTIP}$	0.01	0.09	1	0	0	0	0
Control Variables							
Surplus cash holdings	0.69	0.46	1	1	1	0	0
Free cash flow	0.32	0.47	1	1	0	0	0
Excess investing cash flow	0.19	0.39	1	0	0	0	0
Log(market capitalization)	5.29	2.43	11.71	6.96	5.05	3.30	−0.21
Market-to-book	1.61	0.96	6.29	1.88	1.31	1.00	0.44
Net leverage	−0.15	1.54	0.84	0.19	0.05	−0.15	−24.61
Dividend yield	0.04	0.07	1.42	0.05	0.03	0.02	0.00
Negative returns	0.56	0.50	1	1	1	0	0
Negative earnings	0.10	0.30	1	0	0	0	0
Options outstanding	0.76	18.38	47.91	0.04	0.01	0.00	0.00

	Nonrepurchase Sample (n = 665)							p-value for Difference
	Mean	**Std. Dev.**	**Max**	**Q3**	**Median**	**Q1**	**Min**	
Compensation Variables								
EPS	0.60	0.49	1	1	1	0	0	0.01
NDISC	0.09	0.29	1	0	0	0	0	0.01
EPS_{Bonus}	0.18	0.39	1	0	0	0	0	0.01
$NDISC_{Bonus}$	0.18	0.38	1	0	0	0	0	0.01
EPS_{Option}	0.50	0.50	1	1	1	0	0	0.01
$NDISC_{Option}$	0.01	0.10	1	0	0	0	0	0.01
EPS_{LTIP}	0.19	0.40	1	0	0	0	0	0.61
$NDISC_{LTIP}$	0.01	0.08	1	0	0	0	0	0.75
Control Variables								
Surplus cash holdings	0.65	0.48	1	1	1	0	0	0.09
Free cash flow	0.29	0.45	1	1	0	0	0	0.16
Excess investing cash flow	0.14	0.35	1	0	0	0	0	0.02
Log(market capitalization)	4.72	1.91	10.74	6.06	4.73	3.20	−0.88	0.01
Market-to-book	1.63	1.61	29.69	1.69	1.25	1.00	0.41	0.01
Net leverage	0.03	0.88	1.87	0.29	0.15	−0.02	−10.69	0.01
Dividend yield	0.03	0.09	1.46	0.04	0.02	0.00	0.00	0.01
Negative returns	0.61	0.49	1	1	1	0	0	0.01
Negative earnings	0.28	0.45	1	1	0	0	0	0.01
Options outstanding	0.16	0.77	11.24	0.07	0.02	0.01	0.00	0.01

(continued on next page)

All variables are measured at the beginning of the repurchase year. The "p-value for Difference" column reports probability values for two-tailed paired Wilcoxon Sign Rank tests of the difference between repurchase and nonrepurchase firms.

Variable Definitions (with Extel codes in square brackets where applicable):

NDISC = indicator variable taking the value of 1 when insufficient disclosure renders the presence of an EPS target indeterminate, and 0 otherwise;

EPS = indicator variable taking the value of 1 for firms with at least one bonus plan, option plan, or LTIP tied to EPS performance, and 0 otherwise;

$NDISC_{Bonus}$ = indicator variable taking the value of 1 for firms that fail to explicitly disclose whether bonus payments are conditional on EPS performance, and 0 otherwise;

EPS_{Bonus} = indicator variable taking the value of 1 for firms where bonus payments are fully or partially conditional on EPS performance, and 0 otherwise;

$NDISC_{Option}$ = indicator variable taking the value of 1 for firms that fail to explicitly disclose whether option vesting is conditional on EPS performance, and 0 otherwise;

EPS_{Option} = indicator variable taking the value of 1 for firms where option vesting is fully or partially conditional on EPS performance, and 0 otherwise;

$NDISC_{LTIP}$ = indicator variable taking the value of 1 for firms that fail to explicitly disclose whether LTIP rewards are conditional on EPS performance, and 0 otherwise;

EPS_{LTIP} = indicator variable taking the value of 1 for firms where LTIP rewards are fully or partially conditional on EPS performance, and 0 otherwise;

Surplus cash holdings = indicator variable based on the residual from yearly OLS regressions of the natural logarithm of cash holdings (cash and cash equivalents [ex.CurrentAssetsCashAndNearCash + ex.CurrentAssetsCurrentInvestments] scaled by lagged total assets [ex.TotalAssets] net of cash and cash equivalents) on the natural logarithm of market capitalization [ex.MarketCapitalization], operating cash flow [ex.CFOperatingInflows] scaled by lagged total assets, net working capital (noncash current assets [ex.CurrentAssets − ex.CurrentAssetsCashAndNearCash − ex.CurrentAssetsCurrentInvestments] − current liabilities [ex.Creditors + ex.DebtSTLoans] divided by total assets net of cash and cash equivalents), net leverage (total liabilities [ex.DebtLTLoans + ex.DebtSTLoans] net of cash and cash equivalents divided by total assets net of cash and cash equivalents), research and development ([ex.TradingExpResearchAndDevelopment] divided by total revenue [ex.Sales]), the market-to-book ratio (book value of debt [ex.TotalAssets − ex.ShareholdersEquityOwnEquShares − ex. ShareholdersEquityOwnSharePrem − ex.ShareholdersEquityPreferShares − ex.ShareholdersEquityParticipShares] plus the market value of equity divided by total assets), and dividend payout (an indicator variable taking the value of 1 for firms with non-zero ordinary dividends per share [ex.DividendsPerShareNetReported], and 0 otherwise); *Surplus cash holdings* indicator variable takes the value of 1 where the regression residual is positive, and 0 otherwise;

Free cash flow = indicator variable taking the value of 1 for firms with a market-to-book ratio less than the entire Extel annual sample median and net operating cash flow (scaled by lagged total assets) greater than the entire Extel annual sample median, and 0 otherwise;

Excess investing cash flow = indicator variable taking the value of 1 if investing cash flows [ex.CFInvestments] are positive, and 0 otherwise;

Market capitalization = fiscal year-end share price multiplied by the number of shares outstanding;

Market-to-book = book value of debt plus the market value of equity divided by total assets;

Net leverage = total liabilities net of cash and cash equivalents divided by total assets net of cash and cash equivalents;

Dividend yield = ordinary dividends per share divided by share price;

Negative returns = indicator variable taking the value of 1 if 12-month stock returns are less than the market return over the corresponding period, and 0 otherwise;

Negative earnings = indicator variable taking the value of 1 if earnings per share [ex.EPSAsReported] is negative, and 0 otherwise; and

Options outstanding = aggregate number of outstanding options for all employees at the balance sheet date scaled by market capitalization.

---

chase and control samples are apparent for almost all control variables. Consistent with prior research, repurchasers have more surplus cash, lower net leverage, and are larger than their nonrepurchaser counterparts.

Logistic and Tobit Regressions

Coefficient estimates and model summary statistics for conditional logistic regressions relating the probability of a repurchase to the incidence of EPS-based performance conditions are reported in Table 3, Models 1–3. Results for comparable left-censored tobit regressions are reported in Models 4–5. *EPS* in Model 1 equals 1 when at least one plan of any type links executive compensation to EPS, and 0 otherwise. As predicted, the estimated coefficient on *EPS* is positive and significant. EPS performance conditions also represent an economically important driver of repurchase activity; the predicted odds of a repurchase for firms where executive compensation depends on EPS performance are almost twice the odds for firms where payouts are independent of EPS. Note also that the odds ratio for *EPS* in Model 1 is similar to (and in many cases larger than) the odds ratios associated with traditional determinants of repurchase activity such as excess cash flow and low leverage.

Model 1 also reveals that repurchasers are less likely to disclose details of performance conditions used in executive compensation contracts. The estimated coefficient on *NDISC* is positive and highly significant, while the odds ratio is large relative to other variables. We have no predictions concerning the link between repurchase activity and the transparency with which firms disclose details of compensation arrangements. To ensure that nondiscloser cases are not confounding the analysis, we re-estimated Model 1 after removing such cases. Results reported in Model 2 are entirely consistent with those in Model 1. Whatever effect *NDISC* may be capturing, it appears to be distinct from our main prediction.

Of the remaining control variables in Models 1 and 2, coefficient estimates on *Free cash flow, Excess investing cash flow*, *Market-to-book, Net leverage*, and *Negative earnings* are statistically significant and of the predicted sign. In addition, even after controlling for differences in lagged total assets via our matching procedure, repurchasers are characterized by higher market capitalization at the beginning of the repurchase year.

Model 3 in Table 3 provides evidence on the incremental effects of EPS-based bonus, option, and LTIP arrangements. EPS-based bonus and option plans are associated with incremental positive effects of similar magnitude. In contrast, EPS-based LTIPs have no discernable effect on repurchase activity. The absence of an LTIP effect likely reflects the lower incidence of such plans relative to bonus and option plans (see Table 1, Panel A), coupled with the dominance of total shareholder return (TSR) performance conditions in these plans.[8]

Models 4 and 5 in Table 3 are for left-censored tobit regressions modeling the scaled value of repurchases. Findings replicate those for the conditional logistic models. Even after controlling for traditional determinants of repurchases, annual spending on buybacks is positively associated with *EPS* and the marginal effect for this variable is similar to (and in many cases larger than) the effect for more established drivers of repurchase activity. Consistent with the logistic results, while the coefficient on *NDISC* is positive and highly significant in Model 4, exclusion of nondisclosing matched pairs (Model 5) does not affect the *EPS* coefficient estimate. Collectively, results reported

[8] Although many firms employ EPS targets in their LTIPs, these are often combined with (and subordinate to) a TSR condition. For example, awards under KBC Advanced Technologies' share-based LTIP require minimum TSR over a three-year period equal to the median for the FTSE Small Cap Index over the same period, with maximum awards for TSR at or above 75th percentile ranking. Irrespective of TSR performance, awards are conditional on real EPS growth over the performance period (KBC Advanced Technologies plc Annual Report 2002, 15–16).

TABLE 3

Coefficient Estimates and Model Summary Statistics for Conditional Logistic (Left-Censored Tobit) Regressions Relating the Probability (Scaled Value) of a Repurchase to the Incidence of EPS-Based Performance Conditions in Executive Compensation Contracts

Variables	Predicted Sign	Logistic Models			Tobit Models	
		Model 1	Model 2	Model 3	Model 4	Model 5
Surplus cash holdings	(+)	0.20	0.24	0.24	0.01	0.02
		1.22	*1.27*	*1.27*	*0.00*	*0.00*
		(0.18)	(0.17)	(0.12)	(0.01)	(0.02)
Free cash flow	(+)	0.69	0.60	0.66	0.00	0.00
		2.00	*1.83*	*1.93*	*0.00*	*0.00*
		(0.01)	(0.01)	(0.01)	(0.61)	(0.77)
Excess investing cash flow	(+)	0.71	0.86	0.75	0.02	0.03
		2.03	*2.40*	*2.11*	*0.01*	*0.01*
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Log(market capitalization)	(?)	1.27	1.51	1.22	0.00	0.00
		3.56	*4.52*	*3.39*	*0.00*	*0.00*
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Market-to-book	(−)	−0.72	−0.81	−0.68	0.01	0.01
		0.49	*0.45*	*0.51*	*0.00*	*0.00*
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Net leverage	(−)	−0.43	−0.46	−0.43	−0.01	−0.01
		0.65	*0.63*	*0.65*	*−0.00*	*−0.00*
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Dividend yield	(?)	1.18	1.34	1.15	0.01	0.00
		3.25	*3.82*	*3.15*	*0.00*	*0.00*
		(0.20)	(0.17)	(0.23)	(0.80)	(0.93)
Negative returns	(+)	−0.05	−0.06	−0.03	0.00	0.00
		0.95	*0.94*	*0.97*	*0.00*	*0.00*
		(0.73)	(0.73)	(0.84)	(0.99)	(0.98)
Negative earnings	(−)	−1.22	−1.38	−1.14	−0.05	−0.06
		0.30	*0.25*	*0.32*	*−0.01*	*−0.01*
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Options outstanding	(+)	−0.021	−0.03	−0.02	−0.00	−0.00
		0.98	*0.98*	*0.98*	*−0.00*	*−0.00*
		(0.10)	(0.13)	(0.13)	(0.04)	(0.07)
Earnings-based target	(?)	−0.15	−0.22	−0.06	−0.01	−0.01
		0.86	*0.80*	*0.94*	*−0.00*	*−0.00*
		(0.33)	(0.21)	(0.74)	(0.28)	(0.34)
NDISC	(?)	1.29			0.04	
		3.62			*0.01*	
		(0.01)			(0.01)	
EPS	(+)	0.65	0.71		0.03	0.03
		1.91	*2.03*		*0.01*	*0.01*
		(0.01)	(0.01)		(0.01)	(0.01)
$NDISC_{Bonus}$	(?)			0.54		
				1.72		
				(0.02)		

(continued on next page)

TABLE 3 (continued)

Variables	Predicted Sign	Logistic Models Model 1	Logistic Models Model 2	Logistic Models Model 3	Tobit Models Model 4	Tobit Models Model 5
EPS_{Bonus}	(+)			0.36		
				1.43		
				(0.07)		
$NDISC_{Option}$	(?)			1.95		
				7.03		
				(0.01)		
EPS_{Option}	(+)			0.38		
				1.46		
				(0.02)		
$NDISC_{LTIP}$	(?)			−0.80		
				0.45		
				(0.38)		
EPS_{LTIP}	(+)			−0.08		
				0.92		
				(0.67)		
Intercept	(?)				−0.08	−0.08
					(0.01)	(0.01)
Likelihood ratio		278.44	233.59	283.14	328.63	276.65
p-value		0.01	0.01	0.01		
Pseudo R^2		0.46	0.48	0.46		
% classified correctly		77.30	77.70	75.60		
n		665	524	665	1330	1173
n censored					665	602

Three values are reported for each covariate: the first value is the coefficient estimate; the second (italicized) value is the odds ratio (marginal effect) for logistic (tobit) models; and the third value (in parentheses) is the two-tailed probability value.

The dependent variable in the columns headed "Logistic Models" is the logarithm of the odds of a repurchase. The dependent variable in the columns headed "Tobit Models" is the aggregate amount (including expenses) spent repurchasing shares during fiscal year *t*, scaled by lagged total assets. All variables are measured at the beginning of the repurchase year. Marginal effects for the tobit models are equal to the probability of a no-limit observation multiplied by the coefficient estimate and are evaluated at the mean of each covariate.

Variable Definitions:

NDISC = indicator variable taking the value of 1 when insufficient disclosure renders the presence of an EPS target indeterminate, and 0 otherwise;

EPS = indicator variable taking the value of 1 for firms with at least one of the *j* compensation components (*j* = bonuses, options, or LTIPs) linked to EPS performance, and 0 otherwise;

$NDISC_j$ = indicator variable taking the value of 1 for firms that fail to explicitly disclose whether the *j*th compensation element is conditional on EPS performance, and 0 otherwise;

EPS_j = indicator variable taking the value of 1 for firms where the *j*th compensation element is conditional on EPS performance, and 0 otherwise;

Surplus cash holdings = indicator variable taking the value of 1 where the residual from yearly optimal cash holdings regressions is positive, and 0 otherwise (see Table 2 for details);

Free cash flow = indicator variable taking the value of 1 for firms with a market-to-book ratio less than the sample median for the year and net operating cash flow greater than the sample median for the year, and 0 otherwise;

Excess investing cash flow = indicator variable taking the value of 1 if investing cash flows are positive, and 0 otherwise;

(continued on next page)

TABLE 3 (continued)

Market capitalization = share price multiplied by the number of shares outstanding;
Market-to-book = book value of debt plus the market value of equity divided by total assets;
Net leverage = total liabilities net of cash holdings divided by total assets net of cash holdings;
Dividend yield = ordinary dividends per share divided by share price;
Negative returns = indicator variable taking the value of 1 if 12-month stock returns are less than the market return over the corresponding period, and 0 otherwise;
Negative earnings = indicator variable taking the value of 1 if reported EPS is negative, and 0 otherwise;
Options outstanding = aggregate number of outstanding options for all employees at the balance sheet date scaled by market capitalization; and
Earnings-based target = indicator variable equal to 1 for firms where at least one of the *j* compensation components is linked to any earnings-based metric (including EPS), and 0 otherwise.

in Table 3 provide consistent evidence that EPS-based performance conditions in executive compensation contracts are an important determinant of repurchase activity among U.K. firms.[9]

V. FURTHER ANALYSIS

While tests presented in the previous section establish an association between repurchase activity and performance conditions applied in executive compensation contracts, they leave unresolved the more fundamental question of whether such a link is in shareholders' best interests. This section seeks evidence on the costs and benefits associated with compensation-driven repurchase activity. We begin by testing whether repurchasers with EPS targets exhibit weaker performance (lower payouts to shareholders) as a consequence of management diverting funds from profitable investment opportunities (regular dividends) to fund repurchase activity. Next, we test whether compensation-driven repurchases help to alleviate the problems of surplus cash, suboptimally low leverage, market mispricing, and accrual manipulation.

Post-Repurchase Performance and Payouts

Bens et al. (2002) conclude that U.S. managers divert capital from positive NPV investments to fund repurchases aimed at offsetting the dilutive impact of ESOs. If managers adopt the same strategy to achieve EPS performance targets, then EPS-induced repurchases could represent a serious cost for shareholders in the form of underinvestment. Accordingly, we examine whether a higher probability of repurchases in the presence of EPS-based compensation arrangements is associated with material underinvestment problems.

All else equal, failure to exploit profitable investment opportunities by channeling capital from positive NPV projects to fund repurchases should be reflected in inferior future performance. Tests reported in Table 4 compare the performance of repurchasers and nonrepurchasers using a difference-in-differences OLS specification in which the dependent variable is the change in performance from the pre- to the post-repurchase period. We estimate models for two alternative performance metrics: change in return on assets (ΔROA) and change in market-adjusted stock returns ($\Delta Returns$). The vector of explanatory variables includes indicator variables for repurchas-

[9] In supplementary tests we explored whether the propensity for repurchases is increasing in the number of compensation elements linked explicitly to EPS. The indicator variable EPS_{Single} (EPS_{Multi}) captures firm-years where one (more than one) compensation element is tied to EPS. For conditional logistic regressions, coefficient estimates on EPS_{Single} and EPS_{Multi} are 0.60 and 0.79, respectively, and significant at the $p < 0.01$ level. For tobit regressions, coefficient estimates on both covariates are equal to 0.03 and statistically significant at the $p < 0.01$ level. Results indicate that repurchases are more likely for firms with single and multiple components of pay tied to EPS. However, tests fail to reject equality of coefficient estimates in favor of the alternative hypothesis that the coefficient on $EPS_{Multi} > EPS_{Single}$.

TABLE 4

Coefficient Estimates and Model Summary Statistics for Difference-in-Differences OLS Regressions of Change in Performance Relative to Repurchase Year *t*

	Δ*Performance*: Pre- to Post-Repurchase				**Δ*Performance***	
	Δ*ROA*		**Δ*Returns***		**(*t*−1 to *t*+1) for:**	
Variable	***t*−1 to *t*+1**	***t*−1 to *t*+2**	***t*−1 to *t*+1**	***t*−1 to *t*+2**	**Δ*ROA***	**Δ*Returns***
ROA	−0.32	−0.29	−0.17	−0.04	−0.32	−0.17
	(0.01)	(0.01)	(0.04)	(0.63)	(0.01)	(0.05)
Negative earnings	−0.02	−0.01	0.05	−0.03	−0.02	0.04
	(0.06)	(0.50)	(0.23)	(0.45)	(0.05)	(0.31)
Returns	0.01	0.01	−0.90	−0.94	0.01	−0.90
	(0.15)	(0.09)	(0.01)	(0.01)	(0.14)	(0.01)
Log(Total assets)	0.01	0.01	0.01	−0.01	0.01	0.01
	(0.01)	(0.01)	(0.37)	(0.47)	(0.01)	(0.17)
ΔLog(Total assets)	0.04	0.04	0.20	0.15	0.04	0.19
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Market-to-book	−0.01	−0.01	−0.04	−0.02	−0.01	−0.04
	(0.02)	(0.01)	(0.01)	(0.13)	(0.01)	(0.01)
Leverage	−0.01	−0.04	−0.11	−0.04	−0.01	−0.13
	(0.78)	(0.14)	(0.19)	(0.71)	(0.78)	(0.14)
Working capital	0.00	0.01	−0.01	−0.01	0.00	−0.01
	(0.79)	(0.86)	(0.01)	(0.01)	(0.75)	(0.01)
Investing cash flow	0.02	0.00	−0.02	−0.02	0.03	0.09
	(0.43)	(0.92)	(0.90)	(0.88)	(0.38)	(0.48)
Operating cash flow	0.11	0.03	−0.02	−0.01	0.11	0.05
	(0.01)	(0.48)	(0.87)	(0.94)	(0.01)	(0.75)
Cash holdings	−0.03	−0.06	0.02	−0.24	−0.03	0.04
	(0.35)	(0.07)	(0.88)	(0.06)	(0.39)	(0.73)
Options outstanding	−0.00	−0.00	−0.01	0.01	0.01	−0.01
	(0.39)	(0.42)	(0.01)	(0.94)	(0.61)	(0.01)
Repurchase	−0.01	−0.01	0.11	0.09		
	(0.37)	(0.66)	(0.04)	(0.10)		
EPS	0.01	0.00	0.08	0.00		
	(0.23)	(0.86)	(0.06)	(0.97)		
Repurchase × EPS	0.02	0.01	−0.09	−0.04		
	(0.29)	(0.43)	(0.14)	(0.56)		
$Repurchase_{Cash\ rich}$					−0.04	−0.12
					(0.16)	(0.29)
$Repurchase_{Cash\ constrained}$					−0.02	0.13
					(0.14)	(0.03)
$Nonrepurchase_{Cash\ rich}$					−0.01	−0.01
					(0.49)	(0.86)
$Repurchase_{Cash\ rich} \times EPS$					0.05	0.07
					(0.09)	(0.54)
$Repurchase_{Cash\ constrained} \times EPS$					0.03	−0.10
					(0.03)	(0.12)
$Nonrepurchase_{Cash\ rich} \times EPS$					0.00	0.02
					(0.83)	(0.71)

(continued on next page)

TABLE 4 (continued)

Variable	Δ*Performance*: Pre- to Post-Repurchase				Δ*Performance* (*t*–1 to *t*+1) for:	
	Δ*ROA*		Δ*Returns*			
	t–1 to *t*+1	*t*–1 to *t*+2	*t*–1 to *t*+1	*t*–1 to *t*+2	Δ*ROA*	Δ*Returns*
Intercept	−0.04	−0.02	0.03	0.10	−0.03	0.09
	(0.06)	(0.50)	(0.73)	(0.24)	(0.17)	(0.32)
Industry effects	Yes	Yes	Yes	Yes	Yes	Yes
Year effects	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.21	0.20	0.51	0.55	0.20	0.51
p-value	0.01	0.01	0.01	0.01	0.01	0.01
n	1178	930	1176	930	1178	1176

Performance is measured using return on assets (*ROA*) and 12-month market-adjusted stock returns (*Returns*). Two-tailed probability values are reported in parentheses. All explanatory variables are measured at the beginning of the repurchase year unless otherwise indicated.

Variable Definitions:

Δ*ROA* = change in operating earnings scaled by total assets;
Δ*Returns* = change in 12-month stock returns ending on the fiscal year-end less the equally weighted return on the FTSE All Share index over the corresponding period;
Repurchase = indicator variable equal to 1 for repurchasing firms, and 0 otherwise;
EPS = indicator variable taking the value of 1 for firms with at least one of the *j* compensation components (*j* = bonus plans, option plans, or LTIPs) tied to EPS performance, and 0 otherwise;
$Repurchase_{cash\ rich}$ = indicator variable taking the value of 1 for repurchasing firms where either *Surplus cash holdings, Free cash flow*, or *Excess investing cash flow* from Table 3 is equal to 1 at time *t*–1, and 0 otherwise;
$Repurchase_{Cash\ constrained}$ = indicator variable taking the value of 1 for repurchasing firms where *Surplus cash holdings, Free cash flow*, and *Excess investing cash flow* from Table 3 are equal to 0 at time *t*–1, and 0 otherwise;
$Nonrepurchase_{Cash\ rich}$ = indicator variable taking the value of 1 for nonrepurchasing firms where either *Surplus cash holdings, Free cash flow*, or *Excess investing cash flow* from Table 3 is equal to 1 at time *t*–1, and 0 otherwise;
ROA = return on assets;
Negative earnings = indicator variable taking the value of 1 for firms reporting a loss, and 0 otherwise;
Returns = 12-month market-adjusted stock returns;
Log(total assets) = natural logarithm of total assets;
Δ*Log(total assets)* = change in the natural logarithm of total assets over the same period as the dependent variable is measured;
Market-to-book = book value of debt plus the market value of equity divided by total assets;
Leverage = total liabilities divided by total assets;
Working capital = change in noncash current assets minus liabilities scaled by total assets;
Investing cash flow = cash flow from investing activities scaled by total assets;
Operating cash flow = cash flow from operating activities scaled by total assets;
Cash holdings = cash and cash equivalents scaled by total assets;
Options outstanding = aggregate number of outstanding options for all employees scaled by market capitalization;
Industry effects = vector of industry indicator variables based on Datastream level-3 classification; and
Year effects = vector of calendar year indicator variables.

ers and firms with EPS-based performance conditions, respectively, the associated two-way interaction effect, and a set of controls. Columns 2 and 3 (4 and 5) in Table 4 report results for Δ*ROA* (Δ*Returns*). Findings provide no evidence that repurchasers in general, or repurchases with EPS conditions in particular, exhibit inferior post-repurchase performance. Estimated coefficients on the *Repurchase* indicator variable are either insignificant in the Δ*ROA* models or positive in the

Δ*Returns* models, while coefficient estimates on *Repurchase* $\times$ *EPS* are consistently insignificant. Our findings do not support the view that EPS-based performance targets encourage management to divert cash from positive NPV projects to fund buybacks.

The incentive to channel funds away from profitable investments to support repurchases may be especially strong among the subset of firms with limited cash resources. Columns 6 and 7 in Table 4 extend the previous analysis by replacing *Repurchase* with the following three indicator variables to capture firms' pre-repurchase cash characteristics: $Repurchase_{Cash\ rich}$ takes the value of 1 for repurchasing firms where either *Surplus cash holdings, Free cash flow*, or *Excess investing cash flow* from Table 3 equals 1 at time $t-1$, and 0 otherwise; $Repurchase_{Cash\ constrained}$ takes the value of 1 for repurchasing firms where *Surplus cash holdings, Free cash flow*, and *Excess investing cash flow* equal 0 at time $t-1$, and 0 otherwise; and $Nonrepurchase_{Cash\ rich}$ takes the value of 1 for nonrepurchasing firms where either *Surplus cash holdings, Free cash flow*, or *Excess investing cash flow* equals 1 at time $t-1$, and 0 otherwise. All three indicators are also interacted with *EPS* to capture incremental effects associated with EPS-based compensation. Absent EPS targets, cash-constrained repurchasers are associated with similar accounting performance (column 6) and superior market performance (column 7) in the post-repurchase period relative to other groups. Coefficient estimates on the $Repurchase_{Cash\ constrained} \times EPS$ interaction are either positive (ΔROA model) or indistinguishable from 0 ($\Delta Return$ model). Collectively, these findings provide no evidence that cash-constrained repurchasers, as a group, are characterized by abnormally weak future performance, or that the subset of cash-constrained repurchasers with EPS conditions exhibit incrementally worse performance.

Profitable investment opportunities are not the only source of funds for repurchases that could impose costs on shareholders. A second potentially costly source is regular dividend payments. Because dividends are sticky and dividend cuts are associated with significant stock price declines, regular dividend payments represent a permanent commitment to distribute cash flows and, as such, serve as a disciplinary force on managers' actions (Easterbrook 1984). Repurchases, by comparison, do not imply the same ongoing commitment and are therefore more suited to distributing transitory cash flow shocks (Guay and Harford 2000; Jagannathan et al. 2000). Using cash flows underpinning dividends to fund repurchases could therefore weaken monitoring through a lower commitment to ongoing payouts. Alternatively, repurchases may be associated with increased payouts to shareholders if they enable managers to disgorge lumpy cash surpluses in a timely fashion alongside regular dividend payments.

We examine the impact of EPS-motivated repurchases on payouts to shareholders in two ways. First, we model the annual change in total payouts (dividends plus repurchases scaled by lagged total assets) in the repurchase year using a difference-in-differences tobit regression. Explanatory variables include indicator variables for repurchasers and firms with EPS-based performance conditions, the associated two-way interaction effect, and a vector of controls. Column 2 in Table 5 presents results estimated using the full sample. The *Repurchase* coefficient estimate is positive and significant, indicating that repurchasers experience higher payout increases in the repurchase year. The estimated coefficient on *Repurchase* $\times$ *EPS* is not significant at conventional levels; payout increases for repurchases with EPS-contingent compensation are indistinguishable from those exhibited by non-EPS repurchasers. These results do not support dividend substitution claims for repurchasers in general and EPS-contingent repurchasers in particular. Instead, findings suggest that dividends and repurchases represent complementary payout options that in conjunction yield higher payouts to shareholders regardless of whether repurchases are driven by EPS-contingent compensation arrangements.

Columns 3 and 4 in Table 5 examine change in total payouts for firms with and without surplus cash at time $t-1$, respectively. Results and conclusions are similar to those reported for the full sample in column 2. While results in column 3 provide little evidence that EPS targets are

TABLE 5

Coefficient Estimates and Model Summary Statistics for Pooled Regressions Relating Payout Policy to Repurchase Activity and the Incidence of EPS-Based Performance Conditions in Executive Compensation Contracts

	Tobit Models: Δ*Total Payout*			OLS Model
Variable	**Full Sample**	**Surplus Cash: Yes**	**Surplus Cash: No**	***Dividend Forecast Error***
Surplus cash holdings	0.01			0.03
	(0.06)			(0.24)
Free cash flow	−0.01			0.06
	(0.10)			(0.02)
Excess investing cash flow	−0.01			−0.06
	(0.05)			(0.04)
Log(Market capitalization)	0.00	0.00	0.00	−0.01
	(0.09)	(0.19)	(0.37)	(0.23)
Market-to-book	0.00	0.01	−0.00	0.00
	(0.20)	(0.01)	(0.57)	(0.92)
Net leverage	−0.01	−0.00	−0.01	−0.01
	(0.01)	(0.10)	(0.01)	(0.49)
Negative returns	−0.00	−0.00	0.01	−0.03
	(0.92)	(0.49)	(0.07)	(0.16)
Negative earnings	−0.01	−0.01	−0.03	−0.00
	(0.05)	(0.07)	(0.03)	(0.91)
Options outstanding	0.00	−0.00	0.00	−0.01
	(0.99)	(0.88)	(0.54)	(0.01)
Dividend yield	0.02	−0.00	0.02	0.02
	(0.51)	(0.96)	(0.49)	(0.89)
ROA	0.06	0.06	0.06	−0.03
	(0.01)	(0.01)	(0.08)	(0.66)
Δ*ROA*	−0.03	−0.02	−0.06	−0.08
	(0.05)	(0.07)	(0.17)	(0.22)
Repurchase	0.05	0.05	0.04	−0.04
	(0.01)	(0.01)	(0.01)	(0.18)
Repurchase × EPS	−0.01	−0.01	−0.01	0.06
	(0.30)	(0.36)	(0.72)	(0.10)
Intercept	−0.01	−0.01	−0.02	0.17
	(0.91)	(0.74)	(0.32)	(0.44)
Industry effects	Yes	Yes	Yes	Yes
Year effects	Yes	Yes	Yes	Yes
Likelihood ratio	1370.16	1088.99	301.37	
p-value	0.01	0.01	0.01	0.01
Adjusted R^2				0.04
n	1313	1053	260	1312
n left-censored	214	161	53	

Two-tailed probability values are reported in parentheses.

All explanatory variables are measured at the beginning of the repurchase year unless otherwise indicated.

(continued on next page)

TABLE 5 (continued)

Variable Definitions:

ΔTotal Payout = change in the aggregate value of dividends plus stock repurchases (scaled by lagged total assets) over the period $t-1$ to t; tobit regression models for *ΔTotal Payout* are estimated using three samples: all firms with available data (Full sample), observations where either *Surplus cash holdings*, *Free cash flow*, or *Excess investing cash flow* from Table 3 are equal to 1 at time $t-1$ (Surplus Cash: Yes), and observations where *Surplus cash holdings*, *Free cash flow*, and *Excess investing cash flow* from Table 3 are equal to 0 at time $t-1$ (Surplus Cash: No); and

Dividend Forecast Error = unexpected dividend changes (scaled by lagged market capitalization) over the period $t-1$ to t, computed using a portfolio-based application of Lintner's (1956) model;

Explanatory Variable Definitions:

Repurchase = indicator variable equal to 1 for repurchasing firms, and 0 otherwise;

EPS = indicator variable taking the value of 1 for firms with at least one of the j compensation components (j = bonus plans, option plans, or LTIPs) tied to EPS performance, and 0 otherwise;

ΔROA = change in operating earnings scaled by total assets;

Industry effects = vector of industry indicator variables based on Datastream level-3 classification; and

Year effects = vector of calendar year indicator variables.

See Tables 2 and 3 for definitions of remaining variables.

---

associated with significantly larger payout increases for surplus cash flow repurchasers, there is no suggestion of dividend substitution among low cash flow repurchasers with EPS-contingent compensation (column 4).

Our second test of whether EPS targets encourage managers to divert funds from regular dividend increases to pay for repurchases focuses on dividend forecast errors, defined as the observed dividend change minus the expected dividend change. If repurchasers redirect funds from dividend payments to support repurchase activity, then we should observe lower-than-expected dividend increases in the repurchase year. We draw on Lintner's (1956) model linking current dividends to contemporaneous earnings and lagged dividend payments to estimate expected dividend changes for firm i in year t. Because we are unable to estimate firm-specific regression parameters from Lintner's (1956) model due to insufficient time-series data for earnings and dividends, we instead use a portfolio approach based on the following sequential sort procedure. Each year all nonfinancial firms from Extel with available data are sorted into quintiles based on contemporaneous earnings before exceptional items. Each annual earnings portfolio is then further sorted into quintile portfolios according to lagged ordinary dividends. Median annual dividend changes for the resulting 25 earnings-lagged dividend portfolio combinations (computed after excluding repurchasers) serve as estimates of expected dividend changes. Dividend forecast errors are equal to the observed annual dividend change less the median dividend change for firms' corresponding earnings-lagged dividend portfolio in year t. Column 5 in Table 5 reports coefficient estimates from an OLS regression of dividend forecast errors on the *Repurchase* indicator variable, its corresponding interaction with *EPS*, and a vector of control variables. Although the estimated coefficient on *Repurchase* is negative, as predicted by the dividend-substitution hypothesis, it is not significant at conventional levels ($p = 0.18$). Further, the incremental effect for repurchasers with EPS targets (*Repurchase* × *EPS*) is positive and marginally significant

(p = 0.10). Overall, results do not support the view that EPS-contingent compensation arrangements motivate firms to divert funds from ordinary dividend payments to support repurchases.[10]

Efficient Contracting

Linking executive compensation to EPS growth provides management with a means of manipulating reported performance (through repurchases) that can easily be avoided by using alternative accounting metrics such as ROA or growth in operating profit. Why EPS conditions remain popular in executive compensation contracts despite the additional earnings management opportunities such arrangements provide is, therefore, an intriguing question. One possibility is that EPS conditions represent an efficient contracting device that helps align interests of managers and shareholders by incentivizing managers to make decisions that promote shareholder value. Prior research highlights several shareholder benefits associated with repurchases including limiting overinvestment of surplus cash, increasing leverage in firms with inefficient capital structure, and signaling stock price undervaluation (Dittmar 2000). Insofar as EPS targets create incentives to manipulate reported performance by repurchasing stock, EPS-based compensation may provide shareholders with a simple means of motivating executives to distribute surplus cash, increase leverage, and correct underpricing in a timely manner.

Table 6 presents models testing whether the positive association between repurchases and cash is more pronounced for firms with EPS performance conditions in the presence of surplus cash. We create an indicator variable, *Surplus cash*, taking the value of 1 where *Surplus cash holdings, Free cash flow*, or *Excess investing cash flow* from Table 3 is equal to 1, and 0 otherwise. We use *Surplus cash* to partition the sample before estimating separately for each partition a logistic (left-censored tobit) regression relating the probability (scaled value) of a repurchase to stock and flow measures of cash and their associated interactions with EPS. Estimated coefficients on the *Cash holdings* $\times$ *EPS*, *CFO* $\times$ *EPS*, and *CFI* $\times$ *EPS* interactions are positive and significant in columns 3 and 4, where *Surplus cash* is equal to 1 (except for *Cash holdings* $\times$ *EPS* in the logistic model). In contrast, only the *CFO* $\times$ *EPS* coefficient (tobit model) is positive and significant in columns 5 and 6, where *Surplus cash* is equal to 0. On balance, these findings suggest that EPS targets benefit shareholders by creating a stronger link between repurchases and cash performance in the presence of surplus cash.

Table 7 examines the intervening effect of low leverage and undervaluation on the link between repurchases and EPS targets. In Models 1 and 2, we estimate a version of Equation (1) relating the probability of a repurchase to the presence of EPS targets and an underleverage indicator variable. *Underleverage* takes the value of 1 when net leverage is less than the annual median value for the corresponding Datastream level-4 industry group (computed using the Extel population), and 0 otherwise. Model 1 includes *Underleverage* as a main effect, while Model 2 includes the *Underleverage* $\times$ *EPS* interaction. As predicted, estimated coefficients for the *Underleverage* main effect are positive and significant in both models, as are the coefficients on *EPS*. The *Underleverage* $\times$ *EPS* variable in Model 2, however, does not load (p = 0.93). Accordingly,

[10] In supplementary tests we examined off-diagonal cases based on the predicted probabilities from a pooled version of Model 1 in Table 3. We classified repurchasers with an implied probability $\leq$0.5 as "unexpected repurchasers" and nonrepurchasers with an implied probability >0.5 as "unexpected nonrepurchasers." We compared these cases with observations where the model correctly predicts a repurchase (firm is a repurchaser and the implied probability > 0.5) and a nonrepurchase (firm is a nonrepurchaser and the implied probability $\leq$0.5). We find no evidence that unexpected repurchasers perform worse or make lower payouts than expected repurchasers. These results provide further evidence that EPS-based compensation arrangements do not encourage firms to engage in inconsistently suboptimal repurchase activity. On the other hand, there is some suggestion that unexpected nonrepurchasers perform worse than expected nonrepurchasers, consistent with the view that failure to buy back shares in particular situations imposes agency costs on shareholders (e.g., through overinvestment).

TABLE 6

Coefficient Estimates and Model Summary Statistics for Pooled Logistic (Left-Censored Tobit) Regressions Relating the Probability (Scaled Value) of a Repurchase to the Incidence of EPS-Based Performance Conditions in Executive Compensation Contracts, Conditional on Surplus Cash

		Surplus Cash: Yes		Surplus Cash: No	
Variable	**Predicted Sign**	**Logistic**	**Tobit**	**Logistic**	**Tobit**
Cash holdings	(+)	−3.63	−0.11	0.30	−0.02
		(0.01)	(0.01)	(0.68)	(0.54)
CFO	(+)	3.63	0.14	6.88	0.18
		(0.01)	(0.01)	(0.03)	(0.01)
CFI	(+)	1.34	0.03	6.20	0.14
		(0.01)	(0.04)	(0.09)	(0.08)
Log(Market capitalization)	(?)	0.07	0.00	0.52	0.01
		(0.12)	(0.10)	(0.01)	(0.01)
Market-to-book	(−)	−0.18	0.01	−0.68	−0.01
		(0.10)	(0.01)	(0.03)	(0.48)
Net leverage	(−)	−1.57	−0.02	−0.38	−0.02
		(0.01)	(0.01)	(0.01)	(0.01)
Dividend yield	(?)	−0.67	−0.04	29.80	−0.03
		(0.53)	(0.44)	(0.01)	(0.52)
Negative returns	(+)	−0.05	−0.00	0.40	0.02
		(0.75)	(0.89)	(0.31)	(0.18)
Negative earnings	(−)	−1.14	−0.04	−1.13	−0.02
		(0.01)	(0.01)	(0.12)	(0.32)
Options outstanding	(+)	−0.02	−0.00	0.02	0.01
		(0.06)	(0.35)	(0.77)	(0.47)
NDISC	(?)	1.23	0.03	2.27	0.07
		(0.01)	(0.01)	(0.02)	(0.01)
EPS	(+)	0.45	0.01	−0.35	0.00
		(0.09)	(0.91)	(0.65)	(0.99)
EPS × Cash holdings	(+)	0.50	0.07	−1.39	−0.19
		(0.60)	(0.05)	(0.63)	(0.01)
EPS × CFO	(+)	2.55	0.10	6.00	0.19
		(0.06)	(0.05)	(0.13)	(0.05)
EPS × CFI	(+)	3.42	0.15	1.33	0.04
		(0.01)	(0.01)	(0.76)	(0.71)
Intercept	(?)	−0.02	−0.04	−4.09	−0.09
		(0.96)	(0.01)	(0.01)	(0.01)
Industry		Yes	Yes	Yes	Yes
Year		Yes	Yes	Yes	Yes
Likelihood ratio		233.10	335.66	145.47	90.49
n repurchasers/left-censored		553	516	112	149
Pseudo R^2		0.26		0.57	
% classified correctly		64.50		69.70	
n		1069	1069	261	261

Two tailed probability values are reported in parentheses.

(continued on next page)

TABLE 6 (continued)

The sample is partitioned according to the presence or absence of surplus cash. Columns headed "Surplus Cash: Yes" contain observations where either the *Surplus cash holdings*, *Free cash flow*, or *Excess investing cash flow* indicator variables from Table 3 equals 1. Columns headed "Surplus Cash: No" contain observations where the *Surplus cash holdings*, *Free cash flow*, and *Excess investing cash flow* indicator variables from Table 3 are equal to 0. In columns headed "Logistic" the dependent variable is the logarithm of the odds of a repurchase. In the columns headed "Tobit" the dependent variable is the aggregate amount (including expenses) spent repurchasing shares during fiscal year *t*, scaled by lagged total assets.

All variables are measured at the beginning of the repurchase year.

Explanatory Variable Definitions:

NDISC = indicator variable taking the value of 1 when insufficient disclosure renders the presence of an EPS target indeterminate, and 0 otherwise;
EPS = indicator variable taking the value of 1 for firms with at least one of the *j* compensation components (*j* = bonuses, options, or LTIPs) linked to EPS performance, and 0 otherwise;
Cash holdings = cash and cash equivalents scaled by lagged total assets;
CFO = cash flow from operations scaled by lagged total assets;
CFI = cash flow from investing activities scaled by lagged total assets;
Market capitalization = share price multiplied by the number of shares outstanding;
Market-to-book = book value of debt plus the market value of equity divided by total assets;
Net leverage = total liabilities net of cash holdings divided by total assets net of cash holdings;
Dividend yield = ordinary dividends per share divided by share price;
Negative returns = indicator variable taking the value of 1 if 12-month stock returns are less than the market return over the corresponding period, and 0 otherwise;
Negative earnings = indicator variable taking the value of 1 if reported EPS is negative, and 0 otherwise; and
Options outstanding = aggregate number of outstanding options for all employees at the balance sheet date scaled by market capitalization.

while findings support claims that repurchases are a response to abnormally low leverage, there is no evidence that EPS-based compensation arrangements amplify this effect.

Models 3 and 4 in Table 7 explore the link between equity undervaluation and EPS-motivated repurchases. Our proxy for undervaluation compares observed price with intrinsic value estimated using the harmonic mean of price-to-forward earnings (Liu et al. 2002):

$$IV_{it-1} = E^{F1}_{it-1} \times \frac{1}{\frac{1}{n}\sum_{j=1}^{J}\frac{E^{F1}_{jt-1}}{P_{jt-1}}}, \tag{3}$$

where *IV* is intrinsic value for firm *i* four months after the beginning of repurchase year *t*, E^{F1} is the last available I/B/E/S mean consensus forecast for one-year-ahead EPS during the window -300 days $< t <$ 120 days, *P* is observed stock price four months after the end of fiscal year *t* -1, *J* is the set of firms with consensus forecast data on I/B/E/S in the same Datastream level-3 industry as firm *i* ($i \notin J$), and *n* is the number of firms in *J*. The indicator variable *Undervaluation* takes the value of 1 where $IV_{it-1} > P_{it-1}$, and 0 otherwise. Model 3 in Table 7 includes the *Undervaluation* main effect, while Model 4 is expanded to include the *Undervaluation* × *EPS* term. As with all previous models, *EPS* loads with a statistically significant positive coefficient in both regressions. The *Undervaluation* main effect also loads positively in Model 3, suggesting that repurchases are at least a partial response to market mispricing. However, Model 4 suggests this effect is driven by firms with EPS-based compensation conditions: the coefficient on *Undervaluation* is indistinguishable from 0, whereas *Undervaluation* × *EPS* is positive and significant at the 10 percent level. (The probability value for a likelihood ratio test comparing Models 3 and 4 is 0.09.) An indication of the economic significance of the interaction effect is provided by compar-

TABLE 7

Coefficient Estimates and Model Summary Statistics for Conditional Logistic Regressions Examining the Intervening Effect of Underleverage and Undervaluation on the Association between the Probability of a Repurchase and the Incidence of EPS-Based Performance Conditions in Executive Compensation Contracts

Variables	Predicted Sign	Underleverage Model 1	Underleverage Model 2	Undervaluation Model 3	Undervaluation Model 4
Log(Market capitalization)	(?)	1.21	1.21	1.56	1.56
		3.34	*3.35*	*4.74*	*4.76*
		(0.01)	(0.01)	(0.01)	(0.01)
Market-to-book	(−)	−0.67	−0.68	−0.67	−0.67
		0.51	*0.51*	*0.51*	*0.51*
		(0.01)	(0.01)	(0.01)	(0.01)
Net leverage	(−)			−0.37	−0.41
				0.69	*0.66*
				(0.10)	(0.07)
Dividend yield	(?)	0.90	0.90	0.65	0.65
		2.45	*2.45*	*1.92*	*1.91*
		(0.33)	(0.33)	(0.57)	(0.57)
Negative returns	(+)	−0.04	−0.04		
		0.96	*0.96*		
		(0.78)	(0.78)		
Negative earnings	(−)	−1.08	−1.08	−1.26	−1.23
		0.34	*0.34*	*0.28*	*0.29*
		(0.01)	(0.01)	(0.01)	(0.01)
Options outstanding	(+)	−0.02	−0.02	0.01	0.01
		(0.98)	*(0.98)*	*(1.01)*	*(1.01)*
		(0.12)	(0.12)	(0.75)	(0.71)
Surplus cash holdings	(+)	0.15	0.15	−0.24	−0.25
		1.16	*1.16*	*0.78*	*0.78*
		(0.33)	(0.33)	(0.22)	(0.22)
Free cash flow	(+)	0.63	0.63	0.52	0.53
		1.87	*1.87*	*1.67*	*1.69*
		(0.01)	(0.01)	(0.03)	(0.03)
Excess investing cash flow	(+)	0.72	0.72	0.83	0.83
		2.06	*2.06*	*2.30*	*2.29*
		(0.01)	(0.01)	(0.01)	(0.01)
Underleverage	(+)	0.42	0.40		
		1.52	*1.49*		
		(0.01)	(0.12)		
Undervaluation	(+)			0.76	0.10
				2.13	*1.11*
				(0.01)	(0.83)
NDISC	(?)	1.46	1.46	1.70	1.85
		4.33	*4.33*	*5.46*	*6.36*
		(0.01)	(0.01)	(0.01)	(0.01)
EPS	(+)	0.67	0.65	0.95	0.68
		1.95	*1.92*	*2.57*	*1.98*
		(0.01)	(0.01)	(0.01)	(0.03)

(continued on next page)

TABLE 7 (continued)

Variables	Predicted Sign	Underleverage		Undervaluation	
		Model 1	Model 2	Model 3	Model 4
EPS × Underleverage	(+)		0.03		
			1.03		
			(0.93)		
EPS × Undervaluation	(+)				0.79
					2.20
					(0.10)
Likelihood ratio		267.44	268.17	207.81	210.56
p-value		0.01	0.01	0.01	0.01
Pseudo R^2		0.44	0.44	0.50	0.51
% correctly classified		73.50	73.70	78.10	78.80
n		665	665	439	439

Three values are reported for each covariate: the first value is the coefficient estimate; the second (italicized) value is the odds ratio; and the third value (in parentheses) is the two-tailed probability value.

The dependent variable is the logarithm of the odds of a repurchase. See Tables 2 and 3 for definitions of all remaining variables. All variables are measured at the beginning of the repurchase year.

Variable Definitions:

NDISC = indicator variable taking the value of 1 when insufficient disclosure renders the presence of an EPS target indeterminate, and 0 otherwise;

EPS = indicator variable taking the value of 1 for firms with at least one of the *j* compensation components (*j* = bonus plans, option plans, or LTIPs) tied to EPS performance, and 0 otherwise;

Underleverage = indicator variable taking the value of 1 where net leverage is less than the annual median value for the corresponding Datastream level-4 industry group, and 0 otherwise; and

Undervaluation = indicator variable taking the value of 1 where intrinsic value computed using a price-to-forward-earnings multiple is greater than observed price on the valuation date (four months after the beginning of the repurchase year), and 0 otherwise.

ing the odds ratios for *Undervaluation* in the presence and absence of EPS-contingent contracts.[11] The odds of repurchasing are no different than the odds of not repurchasing for undervalued firms in the absence of EPS-contingent compensation ($e^{0.1} = 1.11$). In contrast, the odds of a repurchase are $e^{0.1+0.79} = 2.44$ times higher than the odds of not repurchasing for undervalued firms with EPS-contingent compensation. Results provide further (albeit statistically weak) evidence that shareholders benefit from EPS-motivated repurchases in the form of higher buyback probability when stock price falls below intrinsic value.

Our final set of tests examines the interaction between stock repurchases and accrual management. Executives need not resort to repurchases to meet binding EPS targets. Prior research indicates that managers often use accounting accruals to maximize compensation payouts (Healy 1985). Although the choice between alternative earnings management instruments has not been widely explored in the literature, several factors suggest that U.K. shareholders favor EPS management through repurchases over accruals. First, because shareholders benefit from repurchases

[11] Direct interpretation of coefficient estimates and odds ratios reported for interaction terms is problematic in nonlinear models since the effect depends on the contribution of the covariates (Ai and Norton 2003). Comparing odds ratios for the *j*th covariate in the presence and absence of the *i*th covariate provides a simple means of interpreting the economic significance of the $j \times i$ interaction.

(see above), the net impact on shareholder wealth from managing EPS through this method is likely to be less detrimental than accruals, where the gains to shareholders are less obvious. Second, the effect of repurchases on reported EPS is more transparent than accrual choices because U.K. firms must disclose details of all repurchase trades to the stock market without delay and report aggregate repurchase activity in their published financial statements. Investors are therefore better placed to reverse the effect of repurchases on reported EPS, should they wish to do so. Third, unlike accruals that reverse over time, buybacks raise the baseline EPS target permanently and do not contribute to future earnings variability.[12]

If managing EPS through repurchases imposes lower net costs on shareholders relative to accrual manipulation, owners would benefit from compensation-driven buybacks via lower accrual manipulation as executives substitute repurchases for discretionary accruals. We test this prediction by regressing measures of absolute abnormal working capital accruals for the repurchase year on an indicator variable for repurchasers with EPS targets (*Repurchase_EPS*), the two-way interaction capturing the incremental effect for repurchasers without EPS conditions (*Repurchase_EPS* $\times$ *NOEPS*), and a vector of controls. We use the absolute value of abnormal accruals because our prediction relates to the overall level of accrual management activity.[13]

Abnormal working capital accruals are estimated using two methods: the modified-Jones model (Dechow et al. 1995) and the performance-matched model (Kothari et al. 2005). We estimate both models cross-sectionally using industry-year portfolios comprising all observations from Extel with available data. Results in Table 8 are consistent with lower abnormal accrual activity for repurchasers with EPS-contingent compensation arrangements, relative to nonrepurchasers and repurchasers without EPS targets. The estimated coefficient on *Repurchase_EPS* is negative and significant in all models, while the *Repurchase_EPS* $\times$ *NOEPS* coefficient is reliably positive and significant. These findings are consistent with the view that firms with EPS targets are more likely to manipulate reported EPS through repurchases as opposed to working capital accruals. To the extent that manipulation via accruals imposes higher costs on shareholders, these results suggest an additional channel through which the positive association between repurchases and EPS-contingent compensation can benefit shareholders.

Summary

Results presented in this section suggest net benefits to shareholders from stock repurchases motivated by EPS targets in executive compensation contracts. Repurchasers are associated with larger increases in total payouts, and this effect is no less pronounced for repurchasers with EPS targets, suggesting that dividend substitution is not a first-order concern. Further, the positive link between repurchases and cash performance is more prominent for firms with EPS targets in the presence of surplus cash; undervalued firms with EPS conditions are more likely to signal their

---

[12] Managers may also favor repurchases over accruals as a means of inflating EPS for several reasons. First, because stock market investors typically view repurchases favorably, they are less likely to question executives' underlying repurchases motives. Second, executives can provide convincing, non-manipulation-based explanations to support their actions if challenged by shareholders.

[13] Using a levels approach to model absolute abnormal accruals introduces the risk of spurious correlation. Our results and conclusions should be interpreted with this caveat in mind. Several factors militate against using a changes specification however. First, accrual reversals confound a changes specification: absolute abnormal accrual levels may remain similar over short intervals even when the level of accrual management declines (because high absolute accruals capture both contemporaneous accrual management and unwinding of accrual management from previous periods). Second, if firms use repurchases instead of accruals in response to a new stimulus for EPS manipulation, then absolute accruals will remain constant over time and a changes specification will yield null results. Consistent with these arguments, when we model the change in absolute accruals we find no difference between repurchasers with EPS conditions, repurchasers without EPS conditions, and nonrepurchasers.

TABLE 8

Coefficient Estimates and Model Summary Statistics from Pooled OLS Regressions Relating Absolute Abnormal Accrual Activity to Repurchase Activity and a Vector of Control Variables

		Absolute Abnormal Accruals Computed Using:			
		DSS Model		KLW Model	
Variable	Predicted Sign	Model 1	Model 2	Model 1	Model 2
Log(Total assets)	(?)		0.00		0.00
			(0.19)		(0.04)
Market-to-book	(?)		0.00		0.00
			(0.29)		(0.14)
Net leverage	(+)		0.01		0.01
			(0.01)		(0.01)
Negative earnings	(+)		0.01		0.01
			(0.14)		(0.52)
Options outstanding	(+)		0.00		0.00
			(0.24)		(0.11)
σOperating cash flow	(+)		0.18		0.20
			(0.01)		(0.01)
Lagged absolute accruals	(+)		0.05		0.05
			(0.01)		(0.01)
Repurchase_EPS	(−)	−0.02	−0.02	−0.02	−0.01
		(0.01)	(0.02)	(0.01)	(0.04)
Repurchase_EPS × NOEPS	(+)	0.02	0.02	0.02	0.02
		(0.05)	(0.05)	(0.02)	(0.05)
Intercept	(?)	0.10	0.09	0.10	0.09
		(0.01)	(0.01)	(0.01)	(0.01)
Industry		No	Yes	No	Yes
Year		No	Yes	No	Yes
Adjusted-R^2		0.01	0.09	0.01	0.07
p-value		0.01	0.01	0.01	0.01
n		1314	1312	1314	1312
n repurchasers		656	654	656	654

Two-tailed probability values are reported in parentheses. The dependent variable is the absolute value of abnormal working capital accruals, where abnormal accruals are the residual from either the Dechow et al. (1995) working capital accrual model (DDS) or the Kothari et al. (2005) working capital accrual model (KLW). Both accrual models are estimated cross-sectionally for all industry-year combinations.

All variables relate to repurchase year t.

Variable Definitions:

Repurchase_EPS = indicator variable equal to 1 for repurchasing firms with at least one of the j compensation components (j = bonus plans, option plans, or LTIPs) tied to EPS performance, and 0 otherwise;

NOEPS = indicator variable taking the value of 1 for firms that do not have at least one of the j compensation components tied to EPS performance, and 0 otherwise;

Total assets = balance sheet value of aggregate assets;

Market-to-book = book value of debt plus the market value of equity divided by total assets;

Net leverage = total liabilities net of cash holdings divided by total assets net of cash holdings;

(continued on next page)

TABLE 8 (continued)

Negative earnings = indicator variable taking the value of 1 if reported earnings per share are negative, and 0 otherwise;
Options outstanding = aggregate number of outstanding options for all employees at the balance sheet date scaled by market capitalization;
σOperating Cash flow = standard deviation of operating cash flow scaled by lagged total assets computed over the three-year period centered on year *t*;
Lagged accruals = one-year lagged value of the dependent variable;
Industry = vector of industry indicator variables based on Datastream level-3 classification; and
Year = vector of calendar year indicator variables.

undervaluation through a repurchase; and repurchasers with EPS conditions are associated with lower abnormal accruals. We find no evidence that EPS-driven repurchases impose costs on shareholders in the form of investment myopia.

VI. CONCLUSIONS

This study examines the impact on firms' stock repurchase activity of EPS performance conditions in executive compensation contracts. Our analysis connects three distinct literatures. One body of research demonstrates how aspects of corporate payout policy are sensitive to executives' compensation arrangements. Another body of work based on surveys and anecdotal evidence indicates that managers are sensitive to the EPS impact of repurchases. A third group of studies concludes that managers use repurchases to achieve key EPS performance thresholds. Our analysis integrates these three literatures by examining how repurchase policy is shaped by contractual arrangements that create an explicit link between executive compensation and reported EPS.

We find that EPS targets explain firm-level repurchase policy after controlling for traditional determinants of buybacks. Further analysis reveals that EPS-motivated repurchases yield net benefits to shareholders. Contrary to Bens et al. (2002), we find no evidence that EPS-driven repurchases lead to investment myopia. Instead, repurchasers are associated with larger payout increases to shareholders that partly reflect a more pronounced link between repurchases and cash performance for firms with EPS targets in the presence of surplus cash flow. In addition, undervalued firms with EPS targets are more likely to signal mispricing through a repurchase, and repurchasers with EPS conditions are associated with lower abnormal accrual activity.

With repurchases emerging as a key payout mechanism in many jurisdictions and in view of concerns about the motives underlying this trend (Bens et al. 2002; Bens et al. 2003; Hribar et al. 2006; Marquardt et al. 2009), a better understanding of the contractual incentives driving repurchase activity and their associated economic consequences is appropriate. Our findings reveal significant contracting benefits from the repurchase incentives that result from linking executive compensation to EPS. In particular, we identify stock repurchases as a potentially important benefit of EPS-based targets in executive compensation contracts that improves alignment of managers' and shareholders' interests. This insight is consistent with Huang et al. (2010) who find that EPS-based bonus plans help to address agency conflicts between managers and shareholders in the form of ownership dilution. We therefore provide a modest further step toward a more complete understanding of the costs and benefits associated with per-share-based performance measures in general, and in particular why EPS-based targets remain a popular choice in executive compensation contracts despite their obvious limitations.

REFERENCES

Aboody, D., and R. Kasznik. 2008. Executive stock-based compensation and firms' cash payout: The role of shareholders' tax-related payout preferences. *Review of Accounting Studies* 13 (2–3): 216–251.

Ai, C., and E. Norton. 2003. Interaction terms in logit and probit models. *Economics Letters* 80 (1): 123–129.

Allison, P. 1999. *Logistic Regression Using the SAS® System: Theory and Application*. Cary, NC: SAS Institute Inc.

Antle, R., and A. Smith. 1986. An empirical investigation of the relative performance evaluation of corporate executives. *Journal of Accounting Research* 24 (1): 1–39.

Badrinath, S., and N. Varaiya. 2000. *The Share Repurchase Decision: Causes, Consequences, and Implementation Guidelines*. Morristown, NJ: Financial Executives Research Foundation.

Bens, D., V. Nagar, D. Skinner, and F. Wong. 2003. Employee stock options, EPS dilution, and stock repurchases. *Journal of Accounting and Economics* 36 (1–3): 51–90.

Bens, D., V. Nagar, and F. Wong. 2002. Real investment implications of employee stock option exercises. *Journal of Accounting Research* 40 (2): 359–393.

Brav, A., J. Graham, and C. Harvey. 2005. Payout policy in the 21st century. *Journal of Financial Economics* 77 (3): 483–527.

Brealey, R., S. Myers, and F. Allen. 2008. *Principles of Corporate Finance*. 9th edition. New York, NY: McGraw-Hill.

Bushman, R., and A. Smith. 2001. Financial accounting information and corporate governance. *Journal of Accounting and Economics* 32 (1–3): 237–333.

Carter, M. E., C. Ittner, and S. Zechman. 2009. Explicit relative performance evaluation in performance-vested equity grants. *Review of Accounting Studies* 14 (2–3): 269–306.

Caster, A. B., R. J. Elson, and L. G. Weld. 2006. Is diluted EPS becoming more art than fact? *The CPA Journal* 76 (9): 26–29.

Cohen, D., A. Dey, and T. Lys. 2008. Real and accrual-based earnings management in the pre- and post-Sarbanes Oxley periods. *The Accounting Review* 83 (3): 757–787.

Conyon, M., and K. Murphy. 2000. The prince and the pauper? CEO pay in the U.S. and U.K. *The Economic Journal* 110 (467): 640–671.

Conyon, M., S. Peck, L. Read, and G. Sadler. 2000. The structure of executive compensation contracts: U.K. evidence. *Long Range Planning* 33 (4): 478–503.

Dechow, P., R. Sloan, and A. Sweeney. 1995. Detecting earnings management. *The Accounting Review* 70 (2): 193–225.

Demski, J., H. Frimor, and D. Sappington. 2004. Efficient manipulation in a repeated setting. *Journal of Accounting Research* 42 (1): 31–49.

Dittmar, A. 2000. Why do firms repurchase stock? *The Journal of Business* 73 (3): 331–355.

Easterbrook, F. 1984. Two agency-cost explanations of dividends. *The American Economic Review* 74 (4): 650–659.

Fenn, G., and N. Liang. 2001. Corporate payout policy and managerial stock incentives. *Journal of Financial Economics* 60 (1): 45–72.

Greenbury Report. 1995. *Directors' Remuneration. Report of a Study Group Chaired by Sir Richard Greenbury*. London, U.K.: Gee Publishing Ltd.

Grullon, G., and R. Michaely. 2002. Dividends, share repurchases, and the substitution hypothesis. *The Journal of Finance* 57 (4): 1649–1684.

Guay, W. 2002. Discussion of: Real investment implications of employee stock option exercises. *Journal of Accounting Research* 40 (2): 395–406.

———, and J. Harford. 2000. The cash flow permanence and information content of dividend increases vs. repurchases. *Journal of Financial Economics* 57 (3): 385–415.

Healy, P. 1985. The effect of bonus schemes on accounting decisions. *Journal of Accounting and Economics* 7 (1–3): 85–107.

Hribar, P., N. Jenkins, and W. B. Johnson. 2006. Stock repurchases as an earnings management device. *Journal of Accounting and Economics* 41 (1–2): 3–27.

Huang, R., C. Marquardt, and B. Zhang. 2010. Why do managers avoid EPS dilution? Presented at the 2010 Journal of Accounting Research Conference; Working paper, City University of New York. Available at SSRN: http://ssrn.com/abstract=1464496.

Jagannathan, M., C. Stephens, and M. Weisbach. 2000. Financial flexibility and the choice between dividends and stock repurchases. *Journal of Financial Economics* 57 (3): 355–384.

Kahle, K. 2002. When a buyback isn't a buyback: Open market repurchases and employee options. *Journal of Financial Economics* 63 (2): 235–261.

Kothari, S. P., A. J. Leone, and C. E. Wasley. 2005. Performance matched discretionary accrual models. *Journal of Accounting and Economics* 39 (1): 163–197.

Lambert, R., W. Lanen, and D. Larcker. 1989. Executive stock option plans and corporate dividend policy. *Journal of Financial and Quantitative Analysis* 24 (4): 409–425.

Lintner, J. 1956. Distribution of incomes of corporations among dividends, retained earnings, and taxes. *The American Economic Review* 46 (2): 97–113.

Liu, J., D. Nissim, and J. Thomas. 2002. Equity valuation using multiples. *Journal of Accounting Research* 40 (1): 135–172.

Marquardt, C., C. Tan, and S. Young. 2009. Accelerated share repurchases, bonus contracts, and CEO horizons. Working paper, City University of New York. Available at: http://ssrn.com/abstract=1346624.

Murphy, K. 1999. Executive compensation. In *Handbook of Labor Economics*, Vol. 3, edited by O. Ashenfelter and D. Card., 2485–2563, Amsterdam, The Netherlands: Elsevier Science North.

Myers, J., L. Myers, and D. Skinner. 2007. Earnings momentum and earnings management. *Journal of Accounting, Auditing & Finance* 22 (2): 249–284.

Opler, T., L. Pinkowitz, R. Stulz, and R. Williamson. 1999. The determinants and implications of corporate cash holdings. *Journal of Financial Economics* 52 (1): 3–46.

———, and S. Titman. 1993. The determinants of leveraged buyout activity: Free cash flows vs. financial distress costs. *The Journal of Finance* 48 (5): 1985–1999.

Pass, C., A. Robinson, and D. Ward. 2000. Performance criteria of corporate option and long-term incentive plans: A survey of 150 U.K. companies 1994–1998. *Management Decision* 38 (2): 130–137.

Prentice, R., and R. Pyke. 1979. Logistic disease incidence models and case-control studies. *Biometrika* 66 (3): 403–411.

Smith, C., and R. Watts. 1992. The investment opportunity set and corporate financing, dividend, and compensation policies. *Journal of Financial Economics* 32 (3): 263–292.

Stephens, C., and M. Weisbach. 1998. Actual share reacquisitions in open market repurchase programs. *The Journal of Finance* 53 (1): 313–333.

Vermaelen, T. 1981. Common stock repurchases and market signaling: An empirical study. *Journal of Financial Economics* 9 (2): 139–183.

Wallace, J. 1997. Adopting residual income-based compensation plans: Do you get what you pay for? *Journal of Accounting and Economics* 24 (3): 275–300.

White, L. 1996. Executive compensation and dividend policy. *Journal of Corporate Finance* 2 (4): 335–358.

THE ACCOUNTING REVIEW
Vol. 86, No. 2
2011
pp. 735–741

American Accounting Association
DOI: 10.2308/accr.00000029

BOOK REVIEWS

Stephen A. Zeff, Editor

Editor's note: Two copies of books for review should be sent to the Book Review Editor: Stephen A. Zeff, Rice University, Jesse H. Jones Graduate School of Business, 6100 Main St., Houston, TX 77005. The policy of *The Accounting Review* is to publish only those reviews solicited by the Book Review Editor. Unsolicited reviews will not be accepted.

DAVID ABOODY and RON KASZNIK, *Executive Compensation and Financial Accounting*, Foundations and Trends® in Accounting (Hanover, MA: now Publishers, Inc., 2009, ISBN: 978-1-60198-342-8, Vol. 4, No. 2, pp. xi, 88).

David Aboody and Ron Kasznik's monograph on the relation between executive compensation and financial reporting explores the interplay of regulation and pay. They organize their review around two themes: (1) compensation provides incentives to influence financial reporting (Chapters 2 and 3) and (2) financial reporting and regulation, more broadly, influence the compensation contracting process (Chapters 4 and 5). The authors do a nice job explaining important institutional details, summarizing much of the relevant literature, and offering potential explanations for conflicting evidence.

The literature in this area is vast and cuts across several disciplines; so no monograph could include all relevant research. And compensation practices and regulations are ever evolving; so new questions continue to arise. Besides highlighting the key features of Aboody and Kasznik's monograph, my aim in this review is to introduce the reader to work that provides different perspectives and highlight additional areas for future research.

I. EFFECTS OF COMPENSATION ON FINANCIAL REPORTING CHOICE

With apologies to the Rolling Stones, this section of the monograph could be subtitled "Maybe you can't always get what you want, but, with a rational economic agent, you just might find you get what you reward." In this section, the authors describe the literature, examining the incentive effects of compensation contracts on managers' financial reporting choices.

In Chapter 2, they summarize the literature on the influence of earnings-based contracts (i.e., bonuses) and stock option compensation on accounting choices. They provide a broad survey of the research in this area and highlight the limitations of studying discretionary accounting choices, including the inability to cleanly measure discretion.

The focus of Chapter 3 is the literature, examining the abnormal (v-shaped) stock price pattern around option grants and, in particular, the influence of stock option grants on voluntary disclosures. The authors primarily discuss their own work, Aboody and Kasznik (2000), which is appropriate, as they were the first researchers to try to disentangle whether managers were timing option grants around corporate disclosures, or vice versa.

The authors provide a survey of the key research in this area and leave the reader with directions for future research. One direction they do not mention is the incentives created by features of other types of equity grants. For example, in the U.K. and increasingly in the U.S., firms are granting stock and stock options with performance-vesting features. Performance requirements can create direct incentives to manipulate earnings or stock price, and reporting requirements make the performance targets more transparent to researchers. But other features of these contracts may alter those incentives (i.e., performance conditions measured over time or proportional vesting over a range of outcomes). Prior research has studied other aspects of these compensation

contracts, but exploring the incentive effects of these other types of equity may be a fruitful direction for future work.[1]

The literature in this section of the monograph demonstrates that compensation contracts can encourage self-serving behavior. Thus, the authors frequently pose the question, "Why would shareholders allow these contracts?" In Section 2.3, they present formally the idea that perhaps creating incentives to make choices is *exactly* in the shareholders' interest and not necessarily rent extraction. They then describe one of their studies, Aboody and Kasznik (2008), as an illustration of contracts motivating what appears to be self-serving behavior but actually serving shareholders' interests. The literature proposes other answers to their question, and thus a broader perspective here would have been helpful. For example, one such answer might be that shareholders are still better off with these contracts, even though they encourage self-serving decisions. Evans and Sridhar (1996) offer insights on this point. Their work considers the trade-off between reporting truthfully and managing earnings, and it recognizes that the firm has two levers at its disposal: the compensation contract that rewards effort, and the financial reporting system that may or may not constrain earnings management. In their model, it can be optimal to offer a compensation contract that may result in earnings management because it would be too costly to eliminate it. Said differently, executives' having skin in the game still provides a greater benefit to shareholders even if, on the margin, executives extract rents.

II. EFFECTS OF REGULATION ON THE DESIGN OF COMPENSATION

The second half of the monograph examines other forces that shape executive compensation—namely, accounting and tax regulation. The authors first take on the question of whether the favorable (i.e., non-expensing) treatment of options prior to SFAS 123R influenced their use. In summarizing that literature, they describe studies that examine the underlying assumption that non-expensing would benefit the firm, possibly through the perception of greater profitability. They then review studies that examine the influence of tax regulation on the design of compensation, focusing on Section 162(m) and various tax treatments of stock options.

The authors provide a nice description of the institutional details of the related accounting and tax issues and draw from a variety of studies in accounting, finance, economics, and management. They also provide critical context for considering the mixed results of the early literature and for assessing the validity of the conclusions of the early literature when compared with later work.

Scholars interested in this area might also consider some additional work. Hodder et al. (2006), for example, provide evidence that discretion allowed in determining option valuation inputs does not necessarily result in Black-Scholes assumptions that are value-decreasing. This study provides an interesting contrast to the studies described in the monograph that suggest that managers use this discretion to lower option expense. Similarly, researchers doing experimental work provide additional evidence that might shape studies examining how the market values stock option expense. For example, Frederickson et al. (2006) provide evidence that financial statement users accord greater reliability to reported over-disclosed expense, while Libby et al. (2006) argue that auditors may require greater accuracy in recognized over-disclosed expense. These findings suggest that the market valuation of stock option compensation may differ under the new accounting treatment. Finally, many studies, too numerous to list, have questioned whether the market values stock options as an expense or as a net benefit. Though diverse in their conclusions, likely from the variety of research designs, these studies may help researchers drill deeper into that question.

As in the prior section of their monograph, the authors lay out many directions for future research, but they may undersell the potential for future work on the relation between tax regulation and pay. While the Tax Code Section 162(m) provides fodder for a stream of research in this area, other tax regulations related to compensation, such as Section 83(b) and Section 409A, may provide interesting additional "frictions." And, unlike the challenges in examining accounting discretion, these studies are able to exploit more "exogenous" events like tax rate changes.

III. GENERAL COMMENTS

The monograph underscores an important message in this line of research, namely, that compensation contracts are shaped by many forces: desire to align incentives, tax rules, financial reporting consequences, resource (cash) constraints, and political costs, among others. All of these forces occur in a labor market where talented executives may demand specific contractual features. This real-world complexity makes research challenging, but

[1] See for example Gerakos et al. (2007), Bettis et al. (2008), and Carter et al. (2009).

we, as scholars, should not be dissuaded by less-than-perfect research designs. A messy world will, at times, yield messy conclusions, but there still might be something we can learn.

I fear that the authors sometimes overlook this reality. In this vein, they are, perhaps, too critical of the backdating research. They note (on page 51) that "inferring backdating in large samples using only publicly available data is challenging. Specifically, it is nearly impossible for researchers … to disentangle backdating from other opportunistic timing of option awards … or opportunistic timing of corporate announcements." This criticism could be leveled at much of the research in this area. And, while undoubtedly imperfect, the results in Lie (2005) were convincing enough to have a hefty impact on regulators and practice.

Similarly, Aboody and Kasznik interpret the mixed results of studies that attempt to document a link between pay and performance as suggesting (on page 2) "weakness in the premise of incentive alignment as a principal determinant of executive compensation." Indeed, this is one of the most fundamental questions of this field, as one solution to the moral hazard problem is the "carrot": deliver better performance, and you shall be rewarded. I would offer, instead, that the mixed results reflect the difficulty of doing work in this area. As researchers, it is rare for us to observe the underlying contract; we only see the payoffs from the contracts. Those payoffs reflect both short-term and long-term objectives of the firm and may rely on performance measures different from accounting or stock price performance often used in these studies. Researchers often must make assumptions about the *ex ante* contract and, to the extent they are wrong, those assumptions weaken the power of their tests.

IV. WHAT NEXT?

At the end of Chapter 5, the authors highlight recent developments, including the Troubled Asset Relief Program, which potentially created frictions in contracting by constraining compensation for participating banks. They also mention the recently enacted "say on pay" legislation, which will increase the visibility of executive compensation. But there are other developments that dovetail with the theme of their monograph. The financial crisis, for example, has engendered real debate about the role that compensation played. Did performance-based pay encourage excessive risk-taking by some firms? Another development is changes to required proxy statement disclosure effective in 2006 and 2010. These new disclosures provide data that allow greater insight into the *ex ante* features of contracts, putting researchers in a better position to examine previously difficult-to-address questions with potentially stronger research designs.

Overall, Aboody and Kasznik offer a coherent and thought-provoking summary of the literature in this area and point to many possibilities for future research. With the continual evolution of compensation practices and related regulation, this is a literature with growth opportunities. To quote Mick Jagger, who, after all, attended the London School of Economics, "A good thing never ends."

REFERENCES

Aboody, D., and R. Kasznik. 2000. CEO stock option awards and the timing of corporate voluntary disclosures. *Journal of Accounting and Economics* 29: 73–100.

———, and ———. 2008. Executive stock-based compensation and firms' cash payout: The role of shareholders' tax-related payout preferences. *Review of Accounting Studies* 13: 216–251.

Bettis, C., J. Bizjak, J. Coles, and S. Kalpathy. 2008. Stock and option grants with performance-based vesting provisions. Working paper, Arizona State University.

Carter, M. E., C. Ittner, and S. Zechman. 2009. Explicit relative performance evaluation in performance-vested equity grants. *Review of Accounting Studies* 14 (2–3): 260–306.

Evans, H., and S. Sridhar. 1996. Multiple control systems, accrual accounting, and earnings management. *Journal of Accounting Research* 34 (1): 45–65.

Frederickson, J., F. Hodge, and J. Pratt. 2006. The evolution of stock option accounting: Disclosure, voluntary recognition, mandated recognition, and management disavowals. *The Accounting Review* 81 (5): 1073–1093.

Gerakos, J., C. Ittner, and D. Larcker. 2007. The structure of performance-vested stock option grants. In *Essays in Accounting Theory in Honour of Joel S. Demski*, Chap. 11, 227–249. New York, NY: Springer Science+Business Media, LLC.

Hodder, L., W. Mayew, M. McAnally, and C. Weaver. 2006. Employee stock option fair-value estimates: Do managerial discretion and incentives explain accuracy? *Contemporary Accounting Research* 23 (4): 933–975.

Libby, R., M. W. Nelson, and J. E. Hunton. 2006. Recognition versus disclosure, auditor tolerance for misstatement, and the reliability of stock-compensation and lease information. *Journal of Accounting Research* 44 (3): 533–560.

Lie, E. 2005. On the timing of CEO stock option awards. *Management Science* 51: 802–812.

MARY ELLEN CARTER
Associate Professor of Accounting
Boston College

JAN PFISTER, *Managing Organizational Culture for Effective Internal Control: From Practice to Theory* (Berlin, Germany: Physica-Verlag, 2009, ISBN 978-3-7908-2339-2340-0, pp. xx, 245).

This book presents the design and outcomes of a field study that aims to develop and validate a new framework for internal control in organizations. It starts with a quick motivation of the importance of such an exercise, which points to the recent (near) collapses of financial institutions across the Western world. Arguing that such failures "can be linked to internal control and organizational culture" (p. 1), the author claims the need to understand the combination of effects of internal control and organizational culture on managerial behavior and organizational outcomes. As, these days, probably few would oppose such a need, the study appears to be a timely academic response to contemporary concerns about the effectiveness of organizational control systems, and potential detrimental effects of organizational culture.

The author's analysis aims to unravel how internal control and organizational culture interact to produce organizational outcomes, guided by an overarching research question that is formulated as follows: "How do management principles and practices affect organizational culture in a way so that organizational culture positively impacts internal control effectiveness?" (p. 6). Whereas the research question proposes a unidirectional model from "management principles and practices," via "organizational culture" to "internal control effectiveness," the author points out elsewhere that the two main challenges of the study lie in capturing "the influence of organizational culture on internal control effectiveness" (p. 3) and "how control mechanisms influence organizational culture" (p. 3), arguing that "not only does culture influence control, but control also influences culture" (p. 3). Disregarding the apparent ambiguity in terminology, the word "challenge" itself seems aptly chosen, because an analysis of the relationships between culture and control is no trivial exercise for theoretical and empirical as well as for conceptual reasons. Chapter 1 therefore starts by presenting definitions of internal control and organizational culture with the aim to delineate them for analytical purposes (pp. 1–2) and continues by pointing out five "broad issues" (p. 5) that the study aims to address. These "broad issues" are problems distilled from the extant literature in management accounting and control, and are presented as "open questions" on the relationship between control and culture that guide the analysis (pp. 5–6).

I find these different layers of proposing the study's structure to be confusing, as it takes some guessing to understand where the author wants our thoughts to go. For example, the first of the "broad issues" relates to the ambiguous nature of culture in causal models of accounting and control practices. The author means that, in extant accounting studies, culture usually acts as an independent variable, even if studies from outside the accounting and control literature conceive culture as a dependent variable. This is quite a straightforward issue, but the author then formulates the open question in rather indirect terms: "how can culture be captured as a construct that is being influenced by control?" (p. 5) A second "broad issue" relates to the distinction between control and culture, as management control typologies sometimes include culture as a type of control. The related "open question" is, "how can culture be captured as a transcendental concept in the context of control?" (p. 6) Three remaining questions point to similar and other ambiguities in the extant literature about the relationship between control and culture and lead to three other open questions. It remains somewhat unclear whether these questions aim to provide the reader with a roadmap to the analysis, or are in fact just illustrations of the logical and semantic complexities of the control and culture debate. In sum, therefore, the brief introduction of "challenges," "broad issues," and a "research question" leaves the reader somewhat confused as to the nature of the theoretical model that will be developed and validated.

The remainder of the introductory chapter is concerned with mapping out the objectives of the study, the

outline of the empirical research design, and an overview of the chapter structure of the book. As regards the objectives, the author emphasizes the aim to contribute to theory as well as practice, pointing out "research gaps" in academic studies and the need for a "deeper knowledge" in practical situations. The research design proposes a qualitative field study that aims to result in an empirically founded framework for the relationship between control and culture. This is why the book is subtitled "from practice to theory."

The remainder of the book is clearly and logically organized into parts and chapters. Part I deals with the "groundwork" of the study and consists of the Chapters 2 and 3. Chapter 2 ("Basics") discusses the various extant definitions and connotations of the concepts of internal control and organizational culture. Since the study is placed in the context of failures of the financial industry (p. 1), it follows that much attention is paid to the effectiveness of internal control (p. 26). Here the author "broadly" (p. 27) defines internal control effectiveness by the extent of legal compliance, reliability of financial statements, and management's assurance that operational effectiveness and efficiency are achieved. It is not entirely clear how this definition supports the analysis. Indeed, while argued to be broad, the definition in fact narrows down internal control to *ex post* assessments via formal systems. While the author seems to suggest on the one hand that management control and internal control are essentially the same, as he refers to typical management control frameworks (e.g., Ouchi 1979; Merchant 1985; Simons 1995), on the other hand he seems to deemphasize the primary focus of management control on good decision making *ex ante*, as well as the fact that these frameworks all include organizational culture. Overall, this left me somewhat confused about the exact definition of internal control that will appear in the final framework. Why does the similarity of internal control and management control receive so much emphasis, after they have been demarcated sharply in Chapter 1 (p. 4)? Why, instead, is no attention paid to the obvious connection between internal control and accounting information systems? Can internal control indeed be seen as including management control, even if it excludes organizational culture? The author continues by arguing that internal control systems, even if "well designed" (p. 30), may not be immune to fraud and errors, and should be defined with a cost-benefit trade-off in mind. It does not become clear, however, how these two characteristics can or should be part of any foundational definition of internal control. The remainder of the second chapter discusses the concept of organizational culture. It leads the reader along various descriptions of organizational culture, resulting in the definition of culture as the ways in which groups of people adapt to the external environment and achieve internal integration (pp. 38–39). Although the definition is intended to exclude formal control arrangements, it does not do so explicitly.

Chapter 3 ("Literature Review") presents a well-structured overview of the conceptual and typological literature on internal control and management accounting and control. After a short section on academic research on internal control, it continues by evaluating the presence or absence of culture in the various internal control frameworks (i.e., COSO, CoCo, Turnbull, etc.). It also assesses the role of culture in management control typologies (e.g., Ouchi 1979; Merchant 1985; Simons 1995) and in a limited number of academic studies on the culture-control relationship. Overall, the findings illustrate the contention expressed earlier that culture is sometimes seen as a "part of" control, and sometimes as an external contingency that affects or interacts with control. The chapter closes with the (correct) conclusion that research on internal control and culture has not yet developed beyond its infancy. In response to this research gap, five "theses" are developed that provide another five guidelines to theory development about the culture-control relationship. An example is the first thesis that "[c]ause and effect between culture and control go in both directions" (p. 65). Taken by themselves, these five theses are a logical conclusion to the chapter. However, in combination with the earlier "research question," the two "challenges," and the five "open questions," a rather complex roadmap emerges that guides the analysis.

Part II, which consists of Chapter 4 ("Field Study"), presents interview feedback from 31 semi-structured interviews the author had with senior officials in 21 companies in Switzerland and in the U.S. (pp. 74–84). Interviews were conducted in two rounds and form the single source of empirical data. A first round of nine interviews in Switzerland aimed at obtaining a "practical understanding of internal control matters in companies" (p. 83). The second round of 22 interviews conducted in the U.S. focused on two questions. First, interviewees from various functions in management, finance, and auditing were asked to identify the principles and practices that they perceive as drivers for control effectiveness. The second question asked the interviewees to identify the roots of control failures. Following a Grounded Theory methodology for content analysis, the author describes the process through which interview transcripts were made, analyzed, grouped, and summarized. This results in the identification of five categories of control failures (related to "commitment," "competence," "communication," "complexity," and "change") and, within each of those categories, five drivers for control effectiveness. These latter drivers describe managerial practices that enhance internal control and mitigate the control problem. For example, practices that mitigate "commitment" failures include "lead by example," "deal with reality," and "ensure accountability." While this framework appears to be plausible, at first sight, the author seems to acknowledge its diversity and lack of clear dimensionality. Therefore, he proposes to interpret the framework as a supportive tool for classification of principles and practices, rather than as the outcome of strict classification. For example, he notes that "the drivers interrelate with each other and often influence several [control failure] areas" (p. 88). The chapter concludes by presenting an overview of illustrative interview quotations for each of the

categories identified, which mention the management practices used to address these failures. Overall, this chapter provides a rich but relatively unstructured overview of internal control challenges and solutions from a selected set of companies and senior officials.

Part III aims to present "theoretical explanations" for the empirical findings from the earlier part. Chapter 5 ("Capturing Culture") sets the scene by presenting culture as a "black-box" that needs to be opened (p. 120). It postulates that we could conceive of culture both as a variable and as a metaphor, and proposes to investigate how culture mediates the relationship between management principles and practices and internal control effectiveness (pp. 120–122). This provides a setup for the analyses in the two subsequent chapters. These analyses distinguish between "the organizational level" (Chapter 6) and "the individual level" (Chapter 7). Chapter 6 relies on social system theory to present culture as an organizational phenomenon, which provides a buffer between the external environment and the internal environment. Via a series of graphical illustrations, the chapter introduces the concepts of "boundaries" between the organization's external environment and its culture, and between its culture and its internal environment. The concept of "transfer" between these three spheres is introduced to explain how behaviors and values are adopted and copied from one sphere to another. It concludes by suggesting that these boundaries may be "open" or "closed," thus enabling or hindering the transfer of behaviors and values between environments and culture. Chapter 7 reiterates this analysis at the individual level, loosely guided by the question of how people's individual interests and values affect the individual's inclination to engage in dysfunctional behavior. It stresses the importance of alignment between individual and organizational values and interests, which is an important, if not a rather undisputed claim.

Part IV ("Synthesis") combines the various steps in the theoretical and empirical analyses into an overall framework. In fact, a surprisingly simple "control-and-culture framework for effective control" is introduced and graphically illustrated (p. 163), which relies on three types of "drivers" that are defined as including "formal and informal control mechanisms, but … also include other factors influencing behavior" (p. 161). The main thrust of this model is that managerial principles and practices provide drivers that may "open" or "close" the "boundaries" between individuals, the culture of the organization they work in, and the external environment in which the organization operates. Closing drivers are those that define how controls need to be performed and how organizational members should act (p. 162). They define acceptable behavior within the culture and provide consistency throughout the culture (p. 165). Opening drivers give those members some freedom to act and adapt (p. 162). They encourage interaction, debate, and accountability (p. 167). A third category of "reinforcing drivers" refers to those that support and communicate the opening and closing drivers that are present (p. 169). On first consideration, this threefold typology is elegant and plausible, as it seems solidly based on the author's account of practice. On second consideration, however, I became a bit concerned with the descriptive and predictive validity of these three kinds of drivers. Can all managerial practices and principles be so neatly ordered to be either of the "opening," "closing," or "reinforcing" type? Are these categories of mechanisms and other factors influencing behavior solid and precise enough to inform theory and practice? Is the difference between "opening," "closing," and "reinforcing" drivers the same at the organizational level as at the individual level, as Chapters 6 and 7 suggest? The remainder of the chapter seems to address these validity concerns by pointing out applications of the new framework. In particular, it gives some examples of how the framework may inform theory and practice. Regarding the former, it shows how the new framework compares to the Merchant and Simons typologies. This is an interesting exercise, even if both extant frameworks are not primarily theory-based or even research-based frameworks. Regarding practice, the author gives some examples of how "opening," "closing," or "reinforcing" drivers (should) reflect in managerial and auditor behavior. The chapter closes by a short review of how the five theses, which were formulated in Chapter 3, can be interpreted in the control-and-culture framework. The chapter furthermore sketches strengths and limitations of the research design. Chapter 9 ("Conclusion"), finally, provides a short overview of the study's questions, design, and outcomes, and concludes with an overview of further research opportunities.

Overall, this study provides a much-needed next step in academic research aimed at understanding the relationships between such complex phenomena as "internal control" and "organizational culture." The choices made in this study regarding theory and method, however, will provide quite a tough read to those expecting an unproblematic analysis in the tradition of contemporary management control studies. Neither contingency theory (dominating behavioral control studies) nor economic theory (dominating agency-based control studies) is explicitly addressed, such that the theoretical foundation of this study appears to be a bit more opaque than in extant research, and to be based on implicit rather than explicit theoretical considerations. Given that the study is empirically founded on only a limited number of observations, especially the author's first suggestion, which is to find generalization of the proposed claims in future studies, deserves support. Indeed, the study provides more than enough theory, empirics, and concepts to make it a starting point for such further enquiries, even if its relative complexity and lack of conceptual, theoretical, and empirical precision does not point to any clear, particular way forward.

REFERENCES

Merchant, K. A. 1985. *Control in Business Organizations*. Boston, MA: Pitman.

Ouchi, W. G. 1979. A conceptual framework for the design of organizational control mechanisms. *Management Science* 25 (9): 833–848.

Simons, R. 1995. *Levers of Control: How Managers Use Innovative Control to Drive Strategic Renewal*. Boston, MA: Harvard Business School Press.

FRANK HARTMANN
Professor of Management Accounting and
Management Control
Erasmus University

EDITORIAL POLICY AND STYLE INFORMATION

EDITORIAL POLICY

According to the policies set by the Publications Committee (which were endorsed by the Executive Committee and were published in the *Accounting Education News*, June 1987), *The Accounting Review* "should be viewed as the premier journal for publishing articles reporting the results of accounting research and explaining and illustrating related research methodology. The scope of acceptable articles should embrace any research methodology and any accounting-related subject, as long as the articles meet the standards established for publication in the journal ... No special sections should be necessary. The primary, but not exclusive, audience should be—as it is now—academicians, graduate students, and others interested in accounting research."

The primary criterion for publication in *The Accounting Review* is the significance of the contribution an article makes to the literature. Topical areas of interest to the journal include accounting information systems, auditing and assurance services, financial accounting, management accounting, taxation, and all other areas of accounting, broadly defined. The journal is also open to all rigorous research methods.

The efficiency and effectiveness of the editorial review process is critically dependent upon the actions of both authors submitting papers and the reviewers. Authors accept the responsibility of preparing research papers at a level suitable for evaluation by independent reviewers. Such preparation, therefore, should include subjecting the manuscript to critique by colleagues and others and revising it accordingly prior to submission. The review process is not to be used as a means of obtaining feedback at early stages of developing the research.

Reviewers and editors are responsible for providing constructive and prompt evaluations of submitted research papers based on the significance of their contribution and on the rigor of analysis and presentation.

MANUSCRIPT PREPARATION AND STYLE

The Accounting Review's manuscript preparation guidelines follow *The Chicago Manual of Style* (15th ed.; University of Chicago Press). Another helpful guide to usage and style is *The Elements of Style*, by William Strunk, Jr., and E. B. White (Macmillan). Spelling follows *Webster's Collegiate Dictionary.*

FORMAT

1. All manuscripts should be formatted in 12-point font on 8 1/2 x 11″ paper and should be double-spaced, except for indented quotations.
2. Manuscripts should be as concise as the subject and research method permit, generally not to exceed 7,000 words.
3. Margins should be at least one inch from top, bottom, and sides.
4. To promote anonymous review, authors should not identify themselves directly or indirectly in their papers or in experimental test instruments included with the submission. Single authors should not use the editorial "we".
5. A cover page should show the title of the paper, all authors' names, titles and affiliations, email addresses, and any acknowledgments.
6. The American Accounting Association encourages use of gender-neutral language in its publications.

Pagination: All pages, including tables, appendices and references, should be serially numbered. Major sections should be numbered in Roman numerals. Subsections should not be numbered.

Numbers: Spell out numbers from one to ten, except when used in tables and lists, and when used with mathematical, statistical, scientific, or technical units and quantities, such as distances, weights and measures. For example: *three days; 3 kilometers; 30 years*. All other numbers are expressed numerically.

Percentages and Decimal Fractions: In nontechnical copy use the word percent in the text.

Hyphens: Use a hyphen to join unit modifiers or to clarify usage. For example: *a cross-sectional equation; re-form*. See *Webster's* for correct usage.

Keywords: The abstract must be followed by at least three keywords to assist in indexing the paper and identifying qualified reviewers.

ABSTRACT/INTRODUCTION

An Abstract of about 100 words (150 maximum) should be presented on a separate page immediately preceeding the text. The Abstract should concisely inform the reader of the manuscript's topic, its methods, and its findings. The Keywords statement should appear immediately below the Abstract. The text of the paper should start with a section labeled "I. Introduction," which provides more details about the paper's purpose, motivation, methodology, and findings. Both the Abstract and the Introduction should be relatively nontechnical, yet clear enough for an informed reader to understand the manuscript's contribution. The manuscript's title, but neither the author's name nor other identification designations, should appear on the Abstract page.

TABLES AND FIGURES

The author should note the following general requirements:

1. Each table and figure (graphic) should appear on a separate page and should be placed at the end of the text. Each should bear an Arabic number and a complete title indicating the exact contents of the table or figure. Tables and figures should define each variable. The titles and definitions should be sufficiently detailed to enable the reader to interpret the tables and figures without reference to the text.
2. A reference to each graphic should be made in the text.
3. The author should indicate where each graphic should be inserted in the text.
4. Graphics should be reasonably interpreted without reference to the text.
5. Source lines and notes should be included as necessary.

Equations: Equations should be numbered in parentheses flush with the right-hand margin.

DOCUMENTATION

Citations: Within-text citations are made using an author-year format. Cited works must correspond to the list of works listed in the "References" section. Authors should make an effort to include the relevant page numbers in the within-text citations.

1. In the text, works are cited as follows: author's last name and year, without comma, in parentheses. For example: one author, (Berry 2003); two authors, (Fehr and Schmidt 2003); three or more authors, (Dechow et al. 1998); more than one work cited, (Cole and Yakushiji 1984; Dechow et al. 1995; Levitt 1998); with two works by the same author(s), (Nelson 2003, 2005).
2. Unless confusion would result, do not use "p." or "pp." before page numbers. For example, (Dechow and Dichev 2002, 41–42).
3. When the reference list contains two or more works by the same author (as the only author or first of two or more authors) published in the same year, the suffix a, b, etc., is appended to the date in the within-text citations and in the "References" section. For example, (Johansson 2004a, 2004b, 2004c; Baiman and Rajan 2002a, 2002b; Dhaliwal et al. 2005a, 2005b).
4. When the author's name is mentioned in the text, it need not be repeated in the citation. For example: "Cohen et al. (2005) provide ..."
5. Citations to institutional works should use acronyms or short titles where practicable. For example: (NCFFR, The Treadway Commission 1987).
6. If the paper refers to statutes, legal treatises, or court cases, citations acceptable in law reviews should be used.

Reference List: Every manuscript must include a "References" section that contains only those works cited within the text. Each entry should contain all information necessary for unambiguous identification of the published work. Use the following formats (which follow *The Chicago Manual of Style*):

1. Arrange citations in *alphabetical order* according to the surname of the first author or the name of the institution or body responsible for the published work.
2. Use authors' initials instead of proper names.
3. For two or more authors, separate authors with a comma, including a comma before "and" (Dechow, P. M., R. Sloan, and A. Sweeney).
4. Date of publication follows the author's (authors') name(s).
5. Titles of journals or newspapers should not be abbreviated.

Sample entries are as follows:

Baiman, S., and M. Rajan. 2002a. The role of information and opportunism in the choice of buyer-supplier relationships. *Journal of Accounting Research* 40 (2): 247–278.

———, and ———. 2002b. Incentive issues in inter-firm relationships. *Accounting, Organizations and Society* 27 (3): 213–238.

Berry, R. 2003. Testimony before the Senate Committee on Homeland Security and Governmental Affairs Permanent Subcommittee on Investigations. November 18. Available at: http://hsgac.senate.gov/files/111803berry.pdf.

Cohen, D., A. Dey, and T. Lys. 2005. The Sarbanes Oxley Act of 2002: Implications for compensation structure and risk-taking incentives of CEOs. Working paper, New York University, University of Chicago, and Northwestern University.

Cole, R., and T. Yakushiji, eds. 1984. *The American and Japanese Auto Industries in Transition.* Ann Arbor, MI: University of Michigan.

Dechow, P. M., R. Sloan, and A. Sweeney. 1995. Detecting earnings management. *The Accounting Review* 70 (2): 193–225.

———, S. P. Kothari, and R. L. Watts. 1998. The relation between earnings and cash flows. *Journal of Accounting and Economics* 25: 133–168.

———, and I. Dichev. 2002. The quality of accruals and earnings: The role of accrual estimation errors, *The Accounting Review* 77 (Supplement): 35–59.

Dhaliwal, D., M. Erickson, and O. Li. 2005a. Shareholder income taxes and the relation between earnings and returns. *Contemporary Accounting Research* 22: 587–616.

———, L. Krull, O. Li, and W. Moser. 2005b. Dividend taxes and implied cost of equity capital. *Journal of Accounting Research* 43: 675–708.

Easton, P. 2003. Discussion of: The predictive value of expenses excluded from pro forma earnings. *Review of Accounting Studies* 8: 175–183.

Engel, E., R. Hayes, and X. Wang. 2007. The Sarbanes-Oxley Act and firms' going-private decisions. *Journal of Accounting and Economics* (forthcoming).

Financial Accounting Standards Board (FASB). 2006. *Accounting for Uncertainty in Income Taxes, and Interpretation of FASB Statement No. 109.* FASB Interpretation No. 48. Financial Accounting Series. Norwalk, CT: FASB.

Fehr, E., and K. Schmidt. 2003. A theory of fairness, competition, and cooperation. In *Advances in Behavioral Economics*, edited by C. Camerer, G. Loewenstein, and M. Rabin, 271–296. New York, NY: Princeton University Press.

Janis, I. L. 1982. *Groupthink: Psychological Studies of Policy Decisions and Fiascoes.* Boston, MA: Houghton Mifflin.

Levitt, A. 1998. The numbers game. Speech delivered at New York University, Center for Law and Business, September 28.

Maggi, G. 1999. The value of commitment with imperfect observability and private information. *RAND Journal of Economics* (Winter) 30: 555–574.

National Commission on Fraudulent Reporting (the Treadway Commission). 1987. *Report of the National Commission on Fraudulent Financial Reporting.* Washington, D.C.: NCFFR.

Nelson, M. W. 2003. Behavioral evidence on the effects of principles- and rules-based standards. *Accounting Horizons* 17 (1): 91–104.

———. 2005. A review of experimental and archival conflicts-of-interest research in auditing. In *Conflicts of Interest: Challenges and Solutions in Business, Law, Medicine, and Public Policy*, edited by D. A. Moore, D. M. Cain, G. Loewenstein, and M. H. Bazerman. Cambridge, U.K.: Cambridge University Press.

Rigdon, E. E., R. E. Schumacker, and W. Wothke. 1998. A comparative review of interaction and nonlinear modeling. In *Interaction and Nonlinear Effects in Structural Equation Modeling*, edited by R. E. Schumacker, and G. A. Marcoulides, 1–16. Mahwah, NJ: Erlbaum Associates.

Schultz, E., and T. Francis. 2002. Companies profit on workers' deaths through "dead peasants" insurance. *Wall Street Journal* (April 19): 1.

Securities and Exchange Commission (SEC). 2002. *Certification of Disclosure in Companies' Quarterly and Annual Reports.* Release Nos. 33-8124, 34-46427. Washington, D.C.: SEC.

U.S. House of Representatives. 2002. The Sarbanes-Oxley Act of 2002. Public Law 107-204 [H. R. 3763]. Washington, D.C.: Government Printing Office.

Footnotes: Footnotes are not used for documentation. Textual footnotes should be used only for extensions and useful excursions of information that if included in the body of the text might disrupt its continuity. Footnotes should be inserted using the "footnote" or "endnote" feature of the word processing software which will automatically number the footnotes throughout the manuscript with superscript Arabic numerals.

SUBMISSION OF MANUSCRIPTS

Authors should note the following guidelines for submitting manuscripts:

1. Manuscripts currently under consideration by another journal or publisher should not be submitted. The author must state upon submission that the work is not submitted or published elsewhere.
2. For manuscripts reporting on field surveys or experiments: If the additional documentation (e.g., questionnaire, case, interview schedule) is sent as a separate file, then all information that might identify the author(s) must be deleted from the instrument.
3. Manuscripts should be submitted via email as Microsoft Word or PDF file to the Senior Editor, Steven Kachelmeier, at email address tar@mccombs.utexas.edu. Please submit separate files for (1) the manuscript's title page with identifying information (not forwarded to reviewers), (2) the manuscript with title page and all other identifying information removed, and (3) any necessary supplemental files, such as experimental instructions and/or response memoranda on invited revisions. The body or subject line of the email should also indicate the submission fee receipt code from AAA (see step 4 below).
4. A nonrefundable submission fee is required unless explicitly waived by the editor for invited revisions of previous submissions. The submission fee in U.S. funds is $200.00 for members and $400.00 for nonmembers of the AAA payable by credit card (VISA or MasterCard only). The payment form is available online at: https://aaahq.org/AAAforms/journals/tarsubmit.cfm. If you are unable to pay by credit card or have any questions, please contact the AAA Member Services Team at (941) 921-7747 or info@aaahq.org.
5. Revisions must be submitted within 12 months from the decision letter inviting a revision.

COMMENTS

The Accounting Review welcomes submissions of comments on previous *TAR* articles. Comments on articles previously published in *The Accounting Review* will generally be reviewed by two reviewers, usually including an author of the original article (to assist the editor in evaluating whether the submitted comment represents the prior article accurately) and an additional reviewer who is independent of the original article. If a comment is accepted for publication, the original author will generally be invited to reply. All other editorial requirements, as enumerated above, apply to proposed comments.

POLICY ON REPRODUCTION

An objective of *The Accounting Review* is to promote the wide dissemination of the results of systematic scholarly inquiries into the broad field of accounting.

Permission is hereby granted to reproduce any of the contents of the *Review* for use in courses of instruction, as long as the source and American Accounting Association copyright are indicated in any such reproductions.

Written application must be made to the American Accounting Association, 5717 Bessie Drive, Sarasota, FL 34233-2399, for permission to reproduce any of the contents of the *Review* for use in other than courses of instruction—e.g., inclusion in books of readings or in any other publications intended for general distribution. In consideration for the grant of permission by the *Review* in such instances, the applicant must notify the author(s) in writing of the intended use to be made of each reproduction.

Except where otherwise noted in articles, the copyright interest has been transferred to the American Accounting Association. Where the author(s) has (have) not transferred the copyright to the Association, applicants must seek permission to reproduce (for all purposes) directly from the author(s).

POLICY ON DATA AVAILABILITY

The AAA's Executive Committee policy (originally adopted in 1989, and amended in 2009) is that the objective of the Association-wide journals (*The Accounting Review, Accounting Horizons, Issues in Accounting Education*) is to provide the widest possible dissemination of knowledge based on systematic scholarly inquiries into accounting as a field of professional research, and educational activity. To fulfill this objective, authors are encouraged to make their data available for use by others in extending or replicating results reported in their articles.

THE ACCOUNTING REVIEW

VOLUME 86, No. 3 MAY 2011

A JOURNAL OF THE
AMERICAN ACCOUNTING ASSOCIATION

AMERICAN ACCOUNTING ASSOCIATION

The American Accounting Association was founded in 1916 as the American Association of University Instructors in Accounting. Its present name was adopted in 1936. The Association is a voluntary organization of persons interested in accounting education, research, and practice.

Membership in the Association is available at the following annual rates:

Full membership:
The membership dues include *Accounting Education News*, and online selection of one, two, or all three of the following Association-wide journals: *The Accounting Review*, *Accounting Horizons*, and *Issues in Accounting Education.* (This does not include Section journals.) Select from:

The Accounting Review only	U.S. $185.00	with Print $220.00
Accounting Horizons OR *Issues in Accounting Education* only	U.S. $185.00	with Print $200.00
The Accounting Review AND *Accounting Horizons* OR *Issues in Accounting Education*	U.S. $195.00	with Print $245.00
Accounting Horizons AND *Issues in Accounting Education*	U.S. $195.00	with Print $225.00
All three Association-wide journals	U.S. $205.00	with Print $270.00

Associate membership:
Full-time students are eligible for associate membership. The print option is not available for Associate members. Select from:

Any one Association-wide journal	U.S. $25.00
Any two Association-wide journals	U.S. $35.00
All three Association-wide journals	U.S. $45.00

In addition, full members may belong to one or more Sections: Accounting, Behavior and Organizations; Accounting Programs Leadership Group; American Taxation Association; Auditing; Diversity; Financial Accounting and Reporting; Forensic and Investigative Accounting; Gender Issues and Worklife Balance; Government and Nonprofit; Information Systems; International Accounting; Management Accounting; Strategic and Emerging Technologies; Teaching, Learning and Curriculum; and Two-Year College. Each Section assesses its own annual dues, which are collected by the Association.

Through May 31, 2011, all editorial correspondence and manuscripts should be emailed to tar@mccombs.utexas.edu. Phone inquiries can be directed to Editorial Assistant Mary Capps at (512) 471-5318; Fax (512) 471-3904. Beginning June 1, 2011, please check online at https://aaahq.org/AAAforms/journals/tarsubmit.cfm for instructions on how to submit a new manuscript to the incoming editor, John Harry Evans III. Revisions of manuscripts originally submitted before June 1, 2011 should be sent via email to the incoming editor, John Harry Evans III, at tar@katz.pitt.edu.

The submission fee of $200.00 for members of the AAA or $400.00 for nonmembers should be paid online (VISA or MasterCard only) at: https://aaahq.org/AAAforms/journals/tarsubmit.cfm and must be paid when article is submitted. Request for reprint permission, subscriptions, advertisements, and dues should be sent to American Accounting Association, 5717 Bessie Drive, Sarasota, FL 34233-2399 • Telephone: (941) 921-7747 • Fax: (941) 923-4093 • Email: Info@aaahq.org. Notice of change of address should be mailed at least eight weeks in advance to ensure undelayed receipt of *The Accounting Review*.

The Accounting Review is indexed in the Social Science Citation Index which is included in Web of Science.

The Accounting Review (ISSN 0001-4826 print and ISSN 1558-7967 online) is published six times a year in January, March, May, July, September, and November by the American Accounting Association, 5717 Bessie Drive, Sarasota, Florida 34233-2399. General subscriptions are available at the annual rates of $395 for a print copy or at $445 for a print and electronic copy. The member subscription rate for a hard copy is $35. Electronic access is included with membership dues. Periodical postage paid at Sarasota, Florida and additional mailing offices. **POSTMASTER:** Send address change to *The Accounting Review*, 5717 Bessie Drive, Sarasota, Florida 34233-2399.

THE ACCOUNTING REVIEW

A JOURNAL OF THE

AMERICAN ACCOUNTING ASSOCIATION

VOL. 86 MAY 2011 No. 3

BOOK REVIEWS, Stephen A. Zeff, Editor

PAST EDITORS

1926–1929	William A. Paton
1929–1943	Eric L. Kohler
1944–1947	A. C. Littleton
1948–1949	Robert L. Dixon
1950–1959	Frank P. Smith
1959–1962	Robert K. Mautz
1962–1965	Lawrence L. Vance
1965–1967	Wendell Trumbull
1968–1970	Charles H. Griffin
1971–1972	Eldon S. Hendriksen
1973–1975	Thomas F. Keller
1976–1978	Don T. DeCoster
1978–1983	Stephen A. Zeff
1983–1987	Gary L. Sundem
1987–1990	William Kinney
1990–1994	A. Rashad Abdel-khalik
1994–1997	Robert P. Magee
1997–2000	Gerald L. Salamon
2000–2002	Linda Smith Bamber
2002–2005	Terry Shevlin
2005–2008	Dan S. Dhaliwal

THE ACCOUNTING REVIEW
Vol. 86, No. 3
2011
pp. 747–767

American Accounting Association
DOI: 10.2308/accr.00000045

Principles-Based versus Rules-Based Accounting Standards: The Influence of Standard Precision and Audit Committee Strength on Financial Reporting Decisions

Christopher P. Agoglia
University of Massachusetts Amherst

Timothy S. Doupnik
University of South Carolina

George T. Tsakumis
Drexel University

ABSTRACT: Recent accounting scandals have resulted in regulatory initiatives designed to strengthen audit committee oversight of corporate financial reporting and have led to a concern that U.S. GAAP has become too rules-based. We examine issues related to these initiatives using two experiments. CFOs in our experiments exhibit more agreement and are less likely to report aggressively under a less precise (more principles-based) standard than under a more precise (more rules-based) standard. Our results also indicate that CFOs applying a more precise standard are less likely to report aggressively in the presence of a strong audit committee than a weak audit committee. We find no effect of audit committee strength when the standard is less precise. Finally, we find support for a three-path mediating model examining mechanisms driving the effect of standard precision on aggressive reporting decisions. These results should be of interest to U.S. policymakers as they continue to contemplate a shift to more principles-based accounting standards (e.g., IFRS).

Keywords: *standard precision; rules-based standards; principles-based standards; audit committee; IFRS.*

Data Availability: *Contact the authors.*

We thank Wendy Bailey, Cathy Beaudoin, Bill Dilla, Rick Hatfield, Scott Jackson, Tamara Lambert, Justin Leiby, Steve Perreault, Roger Silvers, Hun-Tong Tan, Brad Tuttle, Xiaochuan Zheng, Laureen Maines (editor), Steve Kachelmeier (senior editor), two anonymous reviewers, and participants at the 2009 AAA Annual Meeting, the 2009 AAA International Section Midyear Conference, and the 2010 Deloitte/University of Kansas Auditing Symposium for their helpful comments and assistance.

Editor's note: Accepted by Laureen Maines.

Submitted: April 2009
Accepted: December 2010
Published Online: April 2011

I. INTRODUCTION

Our study examines the effect of accounting standard precision on financial statement preparers' reporting judgments, as well as the potential role that the audit committee plays in mitigating aggressive financial reporting under differing levels of standard precision.[1] A wave of corporate accounting scandals in recent years has led to a push for regulatory changes. As part of the Congressional plan to reform U.S. financial reporting, the Sarbanes-Oxley Act of 2002 (SOX, U.S. House of Representatives 2002) introduced major regulatory initiatives in an attempt to overhaul financial reporting and corporate governance systems. These initiatives include two remedies aimed at dampening aggressive financial reporting: (1) the potential adoption of principles-based accounting standards and (2) enhancement of public company audit committees.

Concern has been rising within the financial/investing community that U.S. accounting standards have become too "rules-based." With bright-line tests and detailed guidance, U.S. standards have become so precise that many feel they invite opportunistic interpretation by corporate executives. The perception that a significant number of executives have been concerned with meeting the letter of a rule, more so than its spirit, has led to a call to consider a more "principles-based" regime. Accordingly, SOX required the Securities and Exchange Commission (SEC) to conduct a study on the possible adoption of principles-based standards by the U.S. financial reporting system. Recently, the SEC moved the principles-versus-rules debate to the front burner by proposing a roadmap that could lead to the mandatory adoption of the more principles-based International Financial Reporting Standards (IFRS) by U.S. public companies (SEC 2008).[2]

Additionally, the Financial Accounting Standards Board (FASB) issued a proposal for a principles-based approach to U.S. standard-setting and asked interested parties to comment (FASB 2002). Many who responded were supportive of a principles-based model, believing that such an approach would lead to higher quality financial reporting with less opportunity to "exploit the gaps in GAAP" (CalPERS) or use "financial accounting engineering" to get around detailed, rules-based standards (PricewaterhouseCoopers).[3] Conversely, some argued that less specificity in accounting standards would result in an increase in manipulation of financial results (e.g., Intel and former FASB member David Mosso). On the separate issue of inter-firm comparability, a number of letter-writers expressed concern that, with an increased reliance on judgment, implementation of principles-based standards could result in a decrease in comparability across firms (e.g., IBM, Pfizer, Goldman Sachs, BDO Seidman). The ongoing debate over whether and when to move toward more principles-based accounting standards reflects the uncertainty of both the anticipated desirable and undesirable effects of such a paradigm shift. Although these are empirical questions that can be addressed experimentally to inform policy makers prior to making a move to principles-based standards, the academic literature is limited with respect to research regarding these uncertainties (Maines et al. 2003).

As noted, the possible adoption of principles-based financial reporting standards is only one of the regulatory remedies embodied in SOX to improve U.S. financial reporting. Section 301 of SOX also expands the responsibilities of the audit committee and places greater focus on its role in enhancing the quality of financial reporting. Section 407 requires public companies to disclose

---

1 Similar to Hackenbrack and Nelson (1996), we define aggressive financial reporting as the favorable portrayal of a company's financial situation even when that reporting is not clearly indicated by the facts.

2 On February 24, 2010, the SEC issued a statement reaffirming its support for global accounting standards while pushing back the earliest date for adoption of IFRS to 2015 (SEC 2010).

3 For example, several organizations including Computer Sciences Corporation, PricewaterhouseCoopers, and Financial Executive International made comments indicating that a principles-based approach would result in transactions being accounted for according to their economic substance rather than their form (see the FASB website for these and other comment letters: http://www.fasb.org/jsp/FASB/CommentLetter_C/CommentLetterPage?project_id=1125-001).

whether at least one member of the audit committee is a financial expert. Archival studies have shown audit committee strength to have an effect on the financial reporting process in rules-based standards environments (Klein 2002; Bédard et al. 2004). However, evidence on its effect in more principles-oriented environments and on the interactive effect of standard precision and audit committee strength is lacking (Libby et al. 2002).

To explore these issues, we conduct two experiments in which experienced U.S.-based financial statement preparers are placed in a lease classification decision context. We manipulate standard precision between participants, where more precise and less precise standards are based on FASB Statement No. 13 and International Accounting Standard (IAS) 17, respectively. Audit committee strength is manipulated between participants as either strong or weak.

The results of experiment 1 indicate that financial statement preparers are less likely to report aggressively (i.e., they are more likely to capitalize the lease) when applying a less precise (more principles-based) lease classification criterion than when applying a more precise (more rules-based) criterion. This result provides support for the SEC's and FASB's hope that a move toward more principles-based standards could result in better, or less aggressive, financial reporting. Consistent with prior archival research, we also find that preparers asked to apply a more precise standard are less likely to report aggressively in the presence of a strong audit committee than in the presence of a weak audit committee. However, we find that the influence of a strong audit committee diminishes in the principles-based setting. Interestingly, we find significantly less variability among preparers' reporting decisions when a less precise standard is in place. This suggests that, contrary to the concern of some interested parties, the application of more principles-based standards need not result in less comparability than more precise standards. Experiment 2 confirms the findings of experiment 1 and finds support for a three-path mediation model. Specifically, we find that the less precise the standard, the more concerned preparers are about second-guessing and possible costs imposed through regulation and litigation. This, in turn, results in an increased desire to reflect the underlying economics of transactions and events in the financial statements and, ultimately, in less aggressive reporting.

The findings of our study suggest a number of important implications. For example, these findings should be of interest to policy makers as they contemplate whether and when to adopt IFRS. Our results have encouraging implications for U.S. financial reporting for a move to a more principles-based accounting standards model. Also, our study provides insight regarding recent regulatory changes that address the role and composition of audit committees. Our results suggest that, under certain circumstances, principles-based standards can ease the burden on an audit committee (whether strong or weak) to curtail management's aggressive reporting choices.

Section II provides background information and develops our hypotheses. Sections III and IV describe the experiments used to test hypotheses and present the results. Section V provides a summary and offers conclusions, implications, and suggestions for future research.

II. BACKGROUND AND HYPOTHESES

Principles-Based versus Rules-Based Accounting Standards

The Current Environment

Much of the recent interest in IFRS is predicated on the notion that these standards are principles-based, whereas U.S. GAAP is described as (and often criticized for) being rules-based. Rules-based standards typically provide very detailed guidance with bright-line tests. A perceived benefit of more detailed implementation guidance is greater comparability of financial statements across companies (Schipper 2003). On the other hand, it has been suggested that rules-based standards lead to a "show me where it says I can't" attitude, which, in turn, can lead to dysfunc-

tional financial reporting behavior (Weil 2002, 3). Excessively detailed reporting guidance can invite transaction structuring and incentive-consistent standard interpretation to achieve preferred accounting treatments (FASB 2002; Bockus et al. 2003; Nelson 2003).

The FASB's Statement of Financial Accounting Standards (SFAS) No. 13, *Accounting for Leases*, with its detailed list of criteria for lease classification containing several bright-line thresholds, is cited as the "poster child" for rules-based standards (Maines 2007, 360). In developing SFAS No. 13, the FASB had hoped that explicit lease classification rules would eliminate individual judgment, resulting in consistent application of the standard across firms (Shortridge and Myring 2004). However, because bright-line tests were established, companies have been able to structure and interpret lease contracts to avoid capitalization, which tends to present a more favorable picture of a company's overall financial condition (Imhoff and Thomas 1988; Pulliam 1988). As a result, "the explicit rule allows the off balance sheet financing to continue, *and provides justification for the treatment*" (Shortridge and Myring 2004, 3; emphasis added).

Principles-based standards, which provide limited interpretive and implementation guidance, are the perceived solution to problems caused by rules-based standards. Less guidance, in theory, increases the need to apply professional judgment consistent with the intent of the standards. Both the FASB (2002) and the SEC (2003) believe that this will ultimately result in more meaningful and informative financial statements.[4]

Related Research

Several experimental studies have explored auditor judgments in light of specific accounting standards or guidance (e.g., Gibbins et al. 2001; Ng and Tan 2003; Kadous et al. 2003).[5] For example, Ng and Tan (2003) find that, in the presence of a weak audit committee, auditors are more likely to allow aggressive revenue recognition when no guidance/rule governs the transaction than when there is authoritative guidance specifying the appropriate reporting treatment. Trompeter (1994) finds that, when authoritative guidance limits the range of acceptable accounting treatments, audit partners are less influenced by client preferences. In an imprecise standard context, Hackenbrack and Nelson (1996) show that auditors approve reporting decisions that are consistent with their incentives and use the vagueness of the imprecise standard to justify their decisions. While these studies provide some evidence as to how *auditors* react to the existence or specificity of authoritative guidance, they assume, either implicitly or explicitly, that managers will always choose to report aggressively. This line of research does not speak to how standard precision influences the level of aggressiveness in reporting decisions of financial statement *preparers*.

Three studies have employed an experimental approach to examine the influence of standard precision on reporting decisions made by financial statement preparers (Hoffman and Patton 2002; Psaros and Trotman 2004; Jamal and Tan 2010). Hoffman and Patton (2002) hypothesize that less precise standards for governmental financial reporting provide more latitude for federal financial

---

[4] Specifically, the FASB (2002) concludes that adopting a principles-based approach will result in greater judgment, leading to: accounting treatments that conform to the substance of a transaction, improved transparency, enhanced comparability, increased responsiveness to emerging accounting issues, and facilitation of international financial reporting standard convergence.

[5] Cuccia et al. (1995) explore rule precision in a *tax* setting, finding that moving from a vague rule to a more precise rule has no effect on tax practitioners' propensity to take an aggressive position on an individual's return. When a vague standard exists, tax preparers use the vagueness in the standard to justify their aggressive tax position; when a precise standard exists, they use the vagueness in facts to justify their position. However, as the authors point out, results from their study of tax practitioners might not generalize to financial reporting decisions because of differences in penalties and professional responsibilities (Cuccia et al. 1995), as well as in the scope of impact (e.g., aggressive reporting in an individual tax setting affects a few people at most, limiting the financial/social impact, while aggressive financial reporting by publicly traded companies can have a much broader and deeper impact).

officers to reach reporting decisions that are aligned with their incentives. Contrary to expectations, they find that replacing a less precise recognition threshold ("probable") with a more precise threshold ("more likely than not") does not affect the likelihood that these preparers will recognize a contingent loss. They speculate that the lack of an effect could be due to a weak manipulation, as participants did not appear to perceive the new threshold to be more precise than the original.

Using two separate experiments involving a consolidation judgment, Psaros and Trotman (2004) consider whether corporate accountants will justify aggressive judgments by more aggressively interpreting case-specific information. Results of the first experiment, conducted in a "substance-over-form" (less precise) standard setting, suggest that accountants who make more aggressive judgments tend to interpret case-specific information more aggressively. They obtain similar results in the second experiment, which is set in a more precise standard environment. As the researchers did not include both levels of standard precision in one experiment, they are unable to provide direct evidence on the relative impact that the level of standard precision has on reporting behavior. However, by comparing results from the two experiments, they suggest that aggressive reporting is more likely in rules-based settings. Psaros and Trotman (2004) call for further research designed to more directly compare the relative influence of standards of differing precision on preparers' reporting decisions.

Jamal and Tan (2010) examine whether auditor type (e.g., principles-oriented or rules-oriented) affects financial managers' reporting decisions under rules-based and principles-based standards. They find that auditor type has no effect on reporting decisions under a rules-based standard. However, under a principles-based standard, financial managers are less likely to report aggressively when the auditor is principles-oriented. They speculate that improved financial reporting will result only if a move toward more principles-based standards is accompanied by a shift in auditors' mindsets toward being more principles-oriented. This shift would appear likely to happen relatively quickly because auditors have a responsibility to assure that their clients are in compliance with prevailing standards, regardless of standard type. That is, if there is a move toward principle-based standards such as IFRS, a "rules-based" auditor would quickly have to adjust to an environment in which there are no bright-line thresholds on which to fall back. Thus, the cells from their study that are most relevant for purposes of our study are the two "matched" cells (i.e., principles-oriented auditor/principles-based standard and rules-oriented auditor/rules-based standard). Results from these two cells are not inconsistent with the notion that principles-based standards will typically result in less aggressive financial reporting by preparers.

Standard Precision and Aggressive Financial Reporting

Rules-based standards have developed over time, in part, because of demand for them by financial statement preparers and auditors, who believe that such detailed rules shield them from potential criticism for aggressive reporting (Benston et al. 2006). Aggressive reporting is likely to be more difficult to justify with a less precise standard in which a threshold is not explicitly stated (Maines 2007). The costs associated with being perceived to be out of compliance are similar under both more precise and less precise standards. However, the uncertainty surrounding the risk of being perceived to be out of compliance (i.e., second-guessed) is inherently greater without detailed guidance (Nelson et al. 2002).[6] Thus, preparers applying less precise standards could decide not to select their desired accounting treatment because of the increased risk of second-guessing and the attendant costs. Theoretical research in the areas of law and economics supports the idea that less precise rules could result in less aggressive decisions (Calfee and Craswell 1984; Craswell and Calfee 1986). This outcome could be more likely in the post-SOX era because of

[6] This assumes a strong regulatory enforcement environment (e.g., the post-SOX regulatory environment).

required CEO and CFO certification of the financial statements. Following this reasoning, along with a belief that "objectives-oriented standards allow accounting professionals to operationalize accounting treatments in a manner that best fulfills the objective of each standard and thereby best captures the underlying economic reality," the SEC (2003, 11) and the FASB (2002) have indicated a preference for less precise standards.

There is an implicit assumption underlying this belief that, when the constraints imposed by rules-based standards are removed, financial statement preparers will desire to make accounting choices that best reflect economic reality. However, the FASB (2002, 9) acknowledges that a principles-based approach "could lead to abuse, whereby the principles in accounting standards are not applied in good faith consistent with the intent and spirit of the standards." There are a number of environmental factors that could influence whether a principles-based standard regime leads to more meaningful financial reporting or to abuses of the less-precise guidance. For example, the level of regulatory scrutiny of accounting treatment choices, the degree of consensus among regulators about the appropriate treatment, and the ability of rules-based standards to constrain aggressive choices could influence the efficacy of a principles-based regime.

We choose to examine a setting that best represents the concerns/complaints levied by investors and regulators. Specifically, we examine a rules-based standard that has been unable to constrain aggressive behavior, even in a strong post-SOX regulatory environment in which there is concern about aggressive reporting choices.[7] In such a setting, we expect that the level of standard precision will affect the extent to which financial statement preparers are concerned about second-guessing and possible costs imposed through regulation and litigation. Financial statement preparers should be more concerned about second-guessing when standards are less precise (Nelson et al. 2002). This concern will result in an increased desire to reflect the underlying economics of transactions and events in the financial statements. Presumably, if preparers can show that their accounting choices fairly present economic reality, then they will be better able to defend themselves when second-guessed by external parties (e.g., the SEC). Thus, we argue that less precise standards result in an increased concern about second-guessing and related costs, which in turn results in a greater interest in fair presentation. This reasoning leads to the following hypothesis:

H1: Financial statement preparers applying a less precise financial reporting standard will be less likely to make an aggressive financial reporting decision than preparers applying a more precise standard.

The Influence of the Audit Committee on Aggressive Reporting in Principles-Based and Rules-Based Settings

The audit committee represents a key corporate governance mechanism, in that it is viewed as one of the most important monitors of the financial reporting process (Blue Ribbon Committee 1999). However, a spate of high-profile corporate accounting scandals (e.g., Enron, WorldCom, Tyco) brought the effectiveness of audit committees into question. Regulators and investors have called for stronger audit committees to help ensure higher quality financial reporting. Congress responded by enacting SOX, which, among other things, expands the responsibilities of the audit committee and places a greater focus on its role in enhancing the financial reporting process. Under SOX, the audit committee now is responsible for hiring the company's independent auditor, overseeing the work of the auditor, resolving financial reporting differences between management and the auditor, and monitoring internal controls (Lander 2004). Concurrent with the expansion of

---

[7] This also represents a setting in which we would be likely to find a benefit of a principles-based approach, should such a benefit exist. For example, if a shift to principles-based standards is accompanied by a reduction in the level of regulatory scrutiny, the potential abuses the FASB (2002) acknowledges are more likely to occur.

the audit committee's role, there has been a push toward greater independence and expertise of audit committee members. SOX now requires all members of the audit committee to be independent (i.e., not currently part of management). Companies must now also disclose whether at least one member of the audit committee is a financial expert. Further, while there are no strict guidelines regarding the frequency of audit committee meetings (the Blue Ribbon Committee [1999] recommends at least four meetings per year), some have called for audit committees to be more active and diligent in carrying out their oversight duties (DeZoort et al. 2002; Stewart and Munro 2007).

Thus, conventional wisdom is that audit committee independence, expertise, and activity can enhance the monitoring and oversight of management, leading to less aggressive financial reporting choices. Prior research appears consistent with this notion (Bédard et al. 2004). Results of archival studies examining U.S.-listed firms suggest that independent members are less likely to be sanctioned for fraudulent or misleading financial reporting, and are negatively associated with abnormal accruals, aggressive earnings management, and earnings restatements (Abbott et al. 2000; Klein 2002; Abbott et al. 2004; Bédard et al. 2004). Similarly, financial expertise on the audit committee is negatively associated with discretionary accruals and aggressive earnings management and is viewed favorably by capital markets (Xie et al. 2003; Abbott et al. 2004; Bédard et al. 2004; DeFond et al. 2005). Further, the frequency of audit committee meetings is negatively related to financial reporting misstatements, earnings management, and perceived audit risk (Abbott et al. 2000; Stewart and Munro 2007). Thus, in the rules-based U.S. reporting environment, a strong audit committee can serve as a monitoring mechanism that helps mitigate aggressive financial reporting behavior.

However, it is less clear whether the same relationship exists, or is as strong, in environments that employ principles-based accounting standards. Assuming a relatively strong regulatory environment, preparers using principles-based standards are likely to perceive greater risk of regulator sanctions due to the inherent uncertainty of the standard. This uncertainty could lead to an increased desire to reflect the economic substance of a transaction, in turn lessening the audit committee's burden to constrain aggressive reporting choices. Many countries using the less precise IFRS either do not require audit committees (e.g., Russia) or they are fairly recent developments (e.g., the European Union) and, as a consequence, are likely weak relative to U.S. audit committees. Thus, there is no direct evidence regarding the relationship between audit committee strength and aggressive reporting in a principles-based setting, and what indirect evidence there is does not lead to clear conclusions regarding the existence or strength of the relationship (Collier and Gregory 1996; Peasnell et al. 2000; Chen and Cheng 2007). However, there are certain settings under which inferences could be drawn regarding the influence of audit committee strength in principles-based standard environments.

Recall that, in our setting,[8] we expect that preparers applying a more principles-based standard will be more concerned about regulator second-guessing and capturing the economic substance of the transaction than preparers applying our rules-based standard, irrespective of the strength of the audit committee (i.e., the H1 expectation). If this is the case, then preparers applying the less precise standard will be more likely than preparers applying the more precise standard to choose the treatment preferred by the audit committee, thereby lessening the burden on the audit committee to curb aggressive reporting. That is, the less precise standard should dampen pursuit of aggressive financial reporting regardless of audit committee strength, reducing the need for the audit committee to exert its influence regarding this reporting choice. We therefore test the following hypothesis:

---

[8] Our setting is one in which the rules-based standard has been relatively ineffective at curtailing aggressive reporting, even in a strong post-SOX regulatory environment where there is concern about aggressive reporting choices.

H2: The difference in financial reporting judgments made in the presence of a strong audit committee versus a weak audit committee will be greater when a financial reporting standard is more precise than when it is less precise.

III. EXPERIMENT 1

Participants and Design

Participants in experiment 1 are 96 experienced financial statement preparers (55 CFOs, 23 Controllers, and 18 VPs of Finance) with an average of 25.5 years of professional work experience.[9] As our experiment asks participants to assume the role of a company controller faced with a lease classification decision, it was important that we select experienced executives who are likely to play a key role in the financial reporting decisions of their companies.

To test our hypotheses, we conduct an experiment requiring participants to make a lease classification decision in which the two variables of interest (standard precision and audit committee strength) are manipulated at two levels. The *more precise* condition for standard precision includes a bright-line criterion for lease capitalization (lease term is "equal to 75% or more" of expected life) consistent with SFAS No. 13 (FASB 1976). The *less precise* condition includes a vaguer criterion (lease term is "for the major part" of expected life) based on IAS 17 (IASC 1997). In the *strong audit committee* condition, participants are told that all audit committee members are independent with no disclosed relationship with the company and all qualify as financial experts as defined by the SEC, and that the audit committee meets frequently (11 to 12 times per year). For the *weak audit committee* condition, while all committee members qualify as independent, participants are informed that one member has no disclosed prior relationship with the company (two are former officers of the company), one qualifies as a financial expert, and the committee meets infrequently (two to three times per year).[10]

Procedures

We provided participants with a research instrument containing four sections. Section I included guidelines for classifying a lease as either a capital or operating lease. Participants were told only one criterion (relating to the ratio of lease term to estimated economic life) was relevant to the lease classification decision. Specifically, one group of participants was instructed that a lease must be classified as a capital lease if the lease term is "equal to 75% or more" of the

---

9 We mailed instruments to 1,000 individuals identified by the American Institute of Certified Public Accountants (AICPA) as financial executives. We received replies from 106 individuals and 63 were returned as undeliverable. The resulting response rate of 11.31 percent (106 responses divided by 937 delivered) is consistent with prior studies involving CFO/controller participants (e.g., Graham and Harvey 2001; Gibbins et al. 2007; Sanchez et al. 2007). There were ten unusable responses: seven instruments were completed by inappropriately classified individuals such as staff accountants, payroll clerks, and tax accountants (our conclusions remain the same with or without these individuals), and three were returned with no response on the dependent variable. Thus, 96 usable responses remained. Comparisons of early and late responders indicate no significant difference, suggesting that nonresponse bias does not drive our results. With respect to the demographics of our participants (e.g., current or prior experience working at a publicly traded company or as an external auditor, experience with leases), there were no significant differences between conditions and, when included in our analyses, the demographic variables were neither significant nor altered the conclusions we draw.

10 The features present in both our strong and weak audit committee conditions were selected to represent audit committees that can, and do, exist in today's regulatory environment. For example, our weak committee meets the SOX requirement of independence (no members are currently affiliated with the company) and recommendation for expertise (one member qualifies as a financial expert). With respect to independence, there are numerous publicly traded companies with former employees serving on their audit committees (e.g., FedEx, Goldman Sachs, Kohl's, and Sunoco). Similarly, many firms have a sole designated financial expert (Carcello et al. 2006), including H.J. Heinz Co., CSX Corp., Texas Instruments, Inc., and Nike, Inc. Further, as there are no requirements for frequency of meetings, there is much between-firm variation. Consistent with our weak audit committee manipulation, a number of firms' audit committees meet infrequently (i.e., four or fewer meetings; see Huron 2006).

estimated economic life of the leased property (the more precise standard based on SFAS No. 13). Another group was instructed that a lease must be capitalized if the lease term is "for the major part" of the estimated economic life of the leased property (the less precise standard based on IAS 17). Both groups were told that a lease must be classified as an operating lease if it does not meet the capital lease criterion. In addition, participants were provided the following definitions of "lease term" and "bargain renewal option" based on definitions in both SFAS No. 13 and IAS 17:

- *Lease term* is defined as the fixed non-cancelable term of the lease plus all periods covered by bargain renewal options.
- *Bargain renewal options* allow the lessee to renew the lease for a rental sufficiently lower than the fair rental of the property such that exercise of the option appears, at the inception of the lease, to be reasonably assured.

Given the definition of lease term, lease classification requires financial statement preparers to first judge whether any renewal option embedded in the lease is a bargain. Second, preparers must judge whether the lease term, i.e., the number of years in the fixed non-cancelable portion of the lease plus the number of years in any bargain renewal option, meets the threshold established in the standard.

Section II asked participants to assume the role of controller of a hypothetical company that has entered into a lease for new equipment with an estimated economic life of ten years. The lease has a non-cancelable lease period of seven years. At the end of the initial non-cancelable lease period, the agreement provides the company the option to renew the lease for an additional year, with the monthly rental payment set at a rate that allows for some discretion in judging whether the renewal option represents a bargain.[11] Participants applying the "equal to 75% or more" criterion must exercise their judgment in measuring the "lease term" by determining whether the lease renewal option represents a bargain. They can justify an operating lease classification by arguing that the renewal option does not represent a bargain and therefore the lease term (seven years) is only 70 percent of the asset's economic life (ten years). However, if they judge the renewal option to be a bargain, then they should add the bargain renewal option period (one year) to the initial lease period (seven years) and view the lease term as eight years, which is 80 percent of the asset's economic useful life. In this case, the bright-line threshold in the standard should cause them to classify the lease as a capital lease. Participants applying the less precise standard must exercise similar judgment in determining whether the lease renewal option represents a bargain. However, their interpretation of "for the major part" also factors into their lease classification decisions.[12]

Participants were provided a summary of the differential impact of the two accounting treatments on the company's financial statements and key ratios at the end of the first year of the lease. The summary shows that capitalization of the lease generally produces less favorable financial results, including lower EPS and return on assets figures and a larger debt-to-equity ratio. Thus, participants generally should have an incentive to classify the lease as an operating lease. To make this incentive more salient, participants were told that the company is publicly traded and that it

---

[11] Controllers at several *Fortune* 500 companies suggested that setting the renewal option at approximately 90 percent of the fair rental value would be sufficiently ambiguous to require judgment as to whether it represents a bargain (i.e., whether to include it as part of the total lease term). Thus, we set the option at 90 percent to provide an opportunity for the financial incentives of alternative classification to factor into participants' lease classification decisions.

[12] That is, for the renewal option to place a participant applying the less precise standard in a similar professional decision context as those applying the more precise standard, a less precise standard participant must interpret "for the major part" as ≥71 and ≤80 percent of economic life. We, therefore, conduct an analysis comparing those participants in the 71 to 80 percent range to those in the more precise group (i.e., for both groups of participants, the judgment regarding the option results in a different classification). We discuss this in greater detail below.

intends to raise capital in the second year of the lease term through the issuance of both common stock and the sale of bonds. After considering the case information, participants recorded their lease classification decisions and responded to several post-experimental and demographic questions (Sections III and IV of the instrument).

Experiment 1 Results

Our hypotheses are tested using a 2 × 2 ANOVA (standard precision by audit committee strength) with the financial statement preparer's lease classification decision serving as the dependent variable.[13] Due to the directional nature of expectations, all tests of hypotheses are one-tailed. Cell means and ANOVA results are presented in Table 1.[14]

Standard Precision

H1 predicts that financial statement preparers applying a precise standard will be more likely to make an aggressive financial reporting decision (i.e., classify the lease as an operating lease) than preparers applying an imprecise standard. Preparers recorded their lease classification decisions on a ten-point, forced-choice scale (1 = "Definitely classify as an operating lease" and 10 = "Definitely classify as a capital lease"). Table 1 reports results consistent with expectations. Specifically, preparers are more likely to report aggressively when presented with a more precise standard (mean = 4.98) than with a less precise standard (mean = 7.83, $F = 23.36$, $p < 0.001$).

It is important to establish that these findings are not solely a function of our manipulation of precision resulting in essentially different decision contexts for the less and more precise conditions. That is, for participants in the less precise standard condition, the term "for the major part" is subject to interpretation. For a more meaningful comparison, participants in the less precise condition must be placed in a similar decision context to those in the more precise condition. Specifically, participants' interpretations of "for the major part" must place them in a situation in which their professional judgment regarding the renewal option is meaningful (i.e., the judgment regarding the option results in a different lease classification decision).[15] Of the 47 participants in the less precise group, 19 indicated that they interpreted the phrase within this professional judgment range (i.e., as ≥71 percent and ≤80 percent).[16] As a more restrictive test of the effect of

---

[13] As a manipulation check for audit committee strength, participants were asked to assess (on a seven-point scale where 1 = Low and 7 = High) the company's audit committee along three dimensions: independence from management, accounting/financial expertise, and frequency of meetings. t-tests indicate that participants in the strong audit committee condition assessed the committee's independence (mean = 6.29), accounting/financial expertise (mean = 6.41), and frequency of meetings (mean = 6.63) to be significantly higher (all p's < 0.001) than participants in the weak condition (means of 2.93, 2.52, and 2.84 for independence, expertise, and frequency of meetings, respectively), suggesting that our audit committee manipulation was successful. With respect to the standard precision manipulation, 85 percent of participants correctly identified the applicable standard. Our conclusions remain the same if participants who failed this manipulation check are excluded from the analysis.

[14] Levene's test for equality of variances indicates violations of the ANOVA assumption of homogeneity of variances for the lease decision (discussed later). While ANOVA results are typically quite robust to violations of this assumption (Box 1954; Lindman 1974), we also conduct analyses on transformations (i.e., square root, natural log, inverse function, and ranking transformations) of our dependent variable for assurance. Results of these analyses are consistent with the results presented in Table 1.

[15] That is, if a participant in the less precise condition interprets "for the major part" as ≤70 percent, then the fixed term of the lease alone would automatically result in capitalization. If the interpretation is > 80 percent, then the lease would be treated as an operating lease regardless of whether the renewal option is determined to be a bargain. In either case, the participant's professional judgment regarding the renewal option would be irrelevant to the classification of the lease. Thus, we focus this analysis on those participants whose responses were within the range that requires a "professional judgment" similar to that required of participants in the more precise standard condition.

[16] Mean and median responses for the 47 participants in the less precise group are 69.13 and 70 percent, respectively, with 27 participants interpreting "major part" as ≤70 percent, 19 as within the 71 to 80 percent "judgment required" range, and one as > 80 percent.

TABLE 1

Experiment 1: Lease Classification Decisions (Likelihood Scale)

Panel A: Mean (Standard Deviation)[a]

	Audit Committee		
Standard Precision	**Weak**	**Strong**	**Overall**
Less precise	8.14 (2.24) n = 21	7.58 (2.75) n = 26	7.83 (2.52) n = 47
More precise	4.13 (3.08) n = 23	5.73 (3.51) n = 26	4.98 (3.38) n = 49
Overall	6.05 (3.36) n = 44	6.65 (3.25) n = 52	

Panel B: ANOVA Results

	F	**p-value[b]**
Model	8.76	<0.001
Experimental Variables		
Standard Precision	23.36	<0.001
Audit Committee	0.73	0.396
Interaction		
Standard Precision × Audit Committee	3.19	0.039

[a] Participants indicated the likelihood they would classify a lease as either an operating lease or a capital lease using a ten-point scale numbered from 1 ("Definitely classify as an operating lease") to 10 ("Definitely classify as a capital lease").

[b] Where expectations are directional, p-values are based on one-tailed tests.

standard precision, we retest H1 using only these 19 participants (non-tabulated mean = 7.00 for these individuals). A comparison with responses from the more precise group (mean = 4.98) provides additional support for H1 ($p = 0.013$).[17]

To further explore the effect of standard precision, we dichotomize preparers' responses at the midpoint as a measure of the practical significance of their lease classification decisions. Analysis with the dichotomized variable reveals that participants in the more precise standard condition are significantly more likely to take the aggressive position of classifying the lease as an operating lease than those applying a less precise standard. Sixty-one percent of preparers in the more precise condition classified it as an operating lease versus 15 percent in the less precise condition ($\chi^2 = 21.74$, $p < 0.001$; Table 2), providing further support for H1.

Additionally, in accordance with the development of H1, participants appear to interpret the facts of the lease scenario in a manner consistent with their classification decisions. We find a significant correlation between their classification decisions and their interpretation of whether the

[17] Results are similar when the data are parsed to include only those in the less precise condition who interpret "for the major part" as ≥75 percent (i.e., participants whose interpretations are as high as, or higher than, the threshold in the more precise standard).

TABLE 2
Experiment 1
Contingency Analysis of Lease Classification Decision by Standard Precision Condition

Standard Precision	Lease Classification Decision: Operating Lease	Capital Lease	Row Total
Less precise	7	40	47
Row %	14.89%	85.11%	100.00%
More precise	30	19	49
Row %	61.22%	38.78%	100.00%
Column total	37	59	

$\chi^2 = 21.74$ ($p < 0.001$)

Participants' ten-point likelihood responses were dichotomized at the midpoint such that responses of 1 to 5 on the scale are classified as *Operating Lease* and responses of 6 to 10 are classified as *Capital Lease*.

option to renew the lease at the end of the lease term represents a bargain (Pearson $r = 0.580$; $p < 0.001$). That is, participants who indicate they are more likely to capitalize the lease tend to believe more strongly that the renewal option represents a bargain.

The Influence of the Audit Committee

H2 predicts that audit committee strength will have a greater effect on aggressive reporting in a rules-based regime (where a strong audit committee could inhibit financial statement preparers' opportunistic application of a standard) than in a more principles-based regime (where greater concern about regulator sanctions could lessen the burden on the audit committee to constrain aggressive reporting). Table 1 reports a significant and directionally consistent interaction between standard precision and audit committee strength ($F = 3.19$, $p = 0.039$), providing support for H2. Further, non-tabulated comparisons within each of the standard precision conditions reveal that audit committee strength influences lease classification decisions under the rules-based regime (means = 5.73 and 4.13 for the strong and weak audit committee conditions, respectively, $p = 0.049$), but not under the principles-based regime (means = 7.58 and 8.14 for the strong and weak audit committee conditions, respectively, $p = 0.441$).

An interesting finding emerges when we compare the mean responses in Table 1 in the more precise standard/strong audit committee condition (mean = 5.73) with those in the two less precise standard conditions (means = 7.58 and 8.14 for the less precise/strong and less precise/weak conditions, with two-tailed p-values of 0.040 and 0.007, respectively). We find that both of the less precise standard conditions lead to less aggressive reporting, regardless of audit committee strength. While it is important to stress the context-specific nature of this result, the shift to a more principles-based standard appears to have a greater dampening effect on aggressive reporting than increasing the strength of the audit committee does in our setting.

Comparability

A concern with moving from more precise (rules-based) to less precise (principles-based) standards is that it could result in reduced financial statement comparability across firms (FASB 2002; Nelson 2003; Schipper 2003; SEC 2003). If this concern has merit, one would expect the variability of lease classification decisions to be greater with a less precise standard than with a more precise standard. Contrary to concerns, Table 1, Panel A reveals that there is *less* variability

(as measured by the standard deviation) in the decisions made by preparers in the less precise standard group than in the more precise group (SD = 2.52 and 3.38, respectively). Levene's test for equality of variances indicates a significant difference between the two groups ($F = 15.397$, $p < 0.001$, two-tailed, non-tabulated). This suggests that application of more principles-based standards need not result in less inter-firm comparability than more precise standards.

IV. EXPERIMENT 2

Concern for Sanctions and Economic Substance: A Three-Path Mediating Model

In experiment 1, we find a significant relationship between standard precision and financial statement preparers' lease classification decisions. Specifically, a less precise (more principles-based) standard results in less aggressive financial reporting (H1). The development of H1 suggests two potential mediators of this relationship. With their bright-line thresholds, rules-based standards can invite transaction structuring and incentive-consistent standard interpretation to achieve preferred accounting treatments (FASB 2002; Bockus et al. 2003; Nelson 2003). Thus, as Maines (2007) suggests, financial statement preparers could prefer rules-based standards because they are believed to reduce second-guessing by regulators as to whether a standard has been applied opportunistically. Aggressive reporting could be more difficult with a principles-based standard in which a threshold is not explicitly stated, as the risk of being perceived to be out of compliance is greater due to the inherent uncertainty of the standard (Nelson et al. 2002). When there is uncertainty regarding the proper accounting, preparers might decide not to select their desired treatment because of the increased risk of second-guessing by regulators, as well as the costs accompanying such sanctions (Calfee and Craswell 1984; Craswell and Calfee 1986). This increased concern for second-guessing would result in a greater desire to reflect the underlying economics of transactions and events in the financial statements. That is, if preparers can show that their reporting decisions fairly represent economic reality, they will be better able to defend themselves when second-guessed by regulators. We propose a three-path (two-mediator) model in which concern about second-guessing and interest in reporting the economic substance of a transaction sequentially mediate the relationship between standard precision and aggressive reporting. Experiment 2 is designed to explore this mediating relationship and tests the following hypothesis:

> **H3:** Concern about second-guessing by regulators and the desire to report the economic substance of the transaction will (sequentially) mediate the relationship between standard precision and financial statement preparers' aggressive financial reporting decisions.

Task and Procedure

Similar to experiment 1, participants were 92 experienced financial statement preparers (i.e., CFOs, Controllers, VPs of Finance) with an average of 24.9 years of professional work experience.[18] Experimental materials were similar to those used in experiment 1. Again, we manipulate standard precision and audit committee strength and ask participants to make a lease classification decision. However, we made minor modifications to the details of the lease agreement in an effort to rule out the possibility that a ceiling effect is responsible for our experiment 1 finding that the burden on the audit committee to curb aggressive reporting eases with a move toward a more principles-based standard (i.e., mean responses for participants in both principles-

[18] We conducted another mailing to 1,000 financial executives, receiving 99 replies with 88 returned as undeliverable for a response rate of 10.9 percent. Seven responses were unusable: five were completed by inappropriately classified individuals (our conclusions remain the same with or without these individuals), and two were returned with no response on the dependent variable, resulting in 92 usable responses. Demographic variables (e.g., experience-related measures, years at current job) were not significantly different between conditions and, when included in our analyses, were neither significant nor altered the conclusions we draw.

based conditions were relatively close to the high, or "capitalize," endpoint of the response scale). The modifications to the lease agreement were designed to allow for greater freedom for upward movement on the scale, such that an effect of audit committee strength would be more easily detectible. Specifically, relative to experiment 1, we reduced the fixed non-cancellable term of the lease to five years. The option to renew the lease was increased to three years, but at a rental payment of 92 percent of the equipment's fair rental value. The rental payment was increased to make it easier for participants to conclude that the renewal option is *not* a bargain and thus justify classifying the lease as an operating lease (i.e., to move participants down the response scale to, hopefully, rule out the possibility of a ceiling effect driving our H1 findings for experiment 1). Further, the changes in the length of the fixed term and option increase the proportion of participants in the principles-based standard condition for whom the lease classification requires judgment. These modifications allow for a wider range of participants' interpretations of "for the major part" (i.e., $\geq$51 percent and $\leq$80 percent) to place them in a situation in which their professional judgment regarding the renewal option is meaningful (i.e., the judgment regarding the option results in a different lease classification decision).

In an effort to attempt to identify the mechanisms driving our findings in experiment 1 (and test H3), we ask about several factors that could influence participants' lease classification decisions. Participants were asked, post-experimentally in experiment 2, to consider how much their decisions were influenced by five potentially influential factors (relative to other potential factors). These factors relate to their desire to: (1) "Report the economic substance of the lease in the financial statements," (2) "Avoid possible second-guessing of my decision by the company's audit committee," (3) "Avoid possible second-guessing of my decision by the company's auditor," (4) "Avoid possible second-guessing of my decision by external watchdogs such as the Securities and Exchange Commission," and (5) "Present the company's financial position and profitability as favorably as the circumstances will allow."

Experiment 2 Results

The 2×2 ANOVA results presented in Table 3 confirm the results of experiment 1. Again, we find a significant main effect of standard precision (H1), with preparers applying a more precise standard reporting more aggressively (mean = 3.41) than those applying a less precise standard (mean = 6.19, $p < 0.001$).[19] Dichotomizing participants' responses, we find that participants in the more precise standard condition are significantly more likely to take the aggressive position of classifying the lease as an operating lease (77.3 percent) than those applying a less precise standard (33.3 percent; $\chi^2 = 17.86$, $p < 0.001$, non-tabulated), consistent with H1. Further, we again report significantly less variability in decisions made by preparers in the less precise standard group than in the more precise group (SD = 1.48 and 2.20, respectively; Levene Statistic = 14.918, $p < 0.001$). This result provides additional evidence that principles-based standards might not reduce inter-firm comparability as some have feared. We again find a significant interaction ($p = 0.012$) between standard precision and audit committee strength (H2). Note also that these results appear to rule out the possibility that a ceiling effect is responsible for our H2 results in experiment 1. That is, although the mean responses in the less precise standard conditions were

[19] All participants in the less precise condition interpreted "for the major part" within the range in which their professional judgment regarding the renewal option is required (i.e., $\geq$51 percent and $\leq$80 percent for experiment 2; overall mean and median responses are 68.92 and 70 percent, respectively). Thus, it is unnecessary to conduct the more restrictive test of H1 we utilized for experiment 1. Further, results are similar when the data are parsed to include only those in the less precise condition who interpret "major part" as $\geq$75 percent (i.e., as high as, or higher than, the threshold in the more precise standard). Eighteen of the 48 participants in the less precise condition fall into this category.

TABLE 3

Experiment 2
Lease Classification Decisions (Likelihood Scale)

Panel A: Mean (Standard Deviation)[a]

	Audit Committee		
Standard Precision	**Weak**	**Strong**	**Overall**
Less precise	6.13 (1.46) n = 23	6.24 (1.54) n = 25	6.19 (1.48) n = 48
More precise	2.50 (1.79) n = 22	4.32 (2.23) n = 22	3.41 (2.20) n = 44
Overall	4.36 (2.44) n = 45	5.34 (2.11) n = 47	

Panel B: ANOVA Results

	F	**p-value[b]**
Model	22.76	<0.001
Experimental Variables		
Standard Precision	56.48	<0.001
Audit Committee	6.81	0.011
Interaction		
Standard Precision × Audit Committee	3.13	0.012

[a] Participants indicated the likelihood they would classify a lease as either an operating lease or a capital lease using a ten-point scale numbered from 1 ("Definitely classify as an operating lease") to 10 ("Definitely classify as a capital lease").

[b] Where expectations are directional, p-values are based on one-tailed tests.

a considerable distance away from the "capital lease" endpoint of the scale (overall mean of 6.19 on a ten-point scale), our finding that the influence of audit committee strength diminishes with a move toward a more principles-based standard still holds.

Table 4 presents means for the regulator second-guessing and economic substance variables used to test H3, as well as three other potential influencing factors. Participants recorded on 11-point scales (0 = "Little influence relative to other factors" and 10 = "Very strong influence relative to other factors") the extent to which their lease classification decisions were influenced by these factors. Consistent with our arguments leading to H3, we find that participants applying a less precise standard are more concerned with avoiding regulator second-guessing when making their lease classification decision than those applying a more precise standard (means = 6.64 and 2.14, respectively, $p < 0.001$). Participants in the less precise standard condition also are more concerned with reporting the economic substance of the lease in the financial statements than participants in the more precise standard condition (means of 7.17 and 4.30, respectively, $p < 0.001$).

H3 proposes a three-path mediation model in which concern about second-guessing by regulators and desire to report the economic substance of a transaction sequentially mediate the rela-

TABLE 4
Experiment 2
Factors Influencing the Lease Classification Decision

	Less Precise Standard			More Precise Standard		
	Weak Audit Committee	Strong Audit Committee	Total	Weak Audit Committee	Strong Audit Committee	Total
Dependent Variable	(n = 23)	(n = 25)	(n = 48)	(n = 22)	(n = 22)	(n = 44)
Regulator Second-Guessing						
Mean	6.67	6.60	6.64	1.95	2.32	2.14
(Std. Dev.)	(2.15)	(1.94)	(2.01)	(0.99)	(1.17)	(1.09)
Economic Substance						
Mean	7.09	7.24	7.17	3.77	4.82	4.30
(Std. Dev.)	(2.04)	(1.74)	(1.87)	(1.97)	(1.84)	(1.96)
Audit Committee Second-Guessing						
Mean	1.83	3.44	2.67	1.77	6.00	3.89
(Std. Dev.)	(0.94)	(1.45)	(1.46)	(1.02)	(1.95)	(2.63)
Auditor Second-Guessing						
Mean	3.65	3.60	3.63	3.68	3.91	3.80
(Std. Dev.)	(1.23)	(1.19)	(1.20)	(1.32)	(1.27)	(1.29)
Favorable Financial Presentation						
Mean	3.50	3.28	3.39	7.32	6.41	6.86
(Std. Dev.)	(2.38)	(2.73)	(2.56)	(2.32)	(1.92)	(2.15)

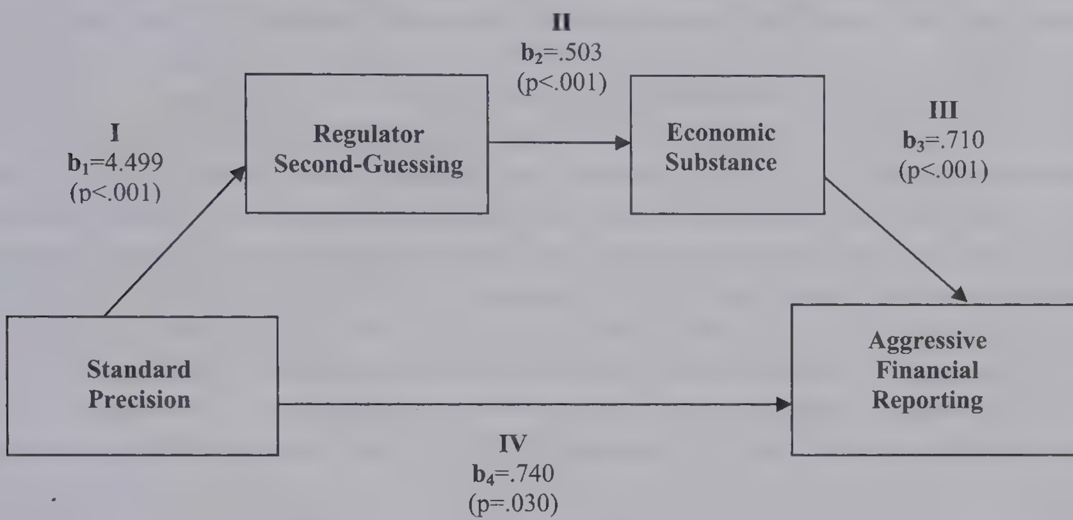
Participants responded to the prompt "Relative to other factors, how much was your lease classification decision influenced by your desire to ..." on 11-point scales, where 0 = "Little influence relative to other factors" and 10 = "Very strong influence relative to other factors" for each of the following factors: "Avoid possible second-guessing of my decision by external watchdogs such as the Securities and Exchange Commission" (*Regulator Second-Guessing*); "Report the economic substance of the lease in the financial statements" (*Economic Substance*); "Avoid possible second-guessing of my decision by the company's audit committee" (*Audit Committee Second-Guessing*); "Avoid possible second-guessing of my decision by the company's auditor" (*Auditor Second-Guessing*); and "Present the company's financial position and profitability as favorably as the circumstances will allow" (*Favorable Financial Position*).

tionship between standard precision and aggressive reporting. Figure 1 reports the results of our test of this model, which support H3. The predicted mediating relationship is demonstrated by the following regression results (MacKinnon and Dwyer 1993):

(1) Standard precision significantly affects concern for regulator second-guessing (path I in Figure 1, $b_1 = 4.499$, $p < 0.001$),
(2) Concern for regulator second-guessing influences desire to present the economic substance of the transaction (path II, $b_2 = 0.503$, $p < 0.001$),
(3) Concern for reporting the economic substance of the transaction significantly influences aggressive financial reporting (path III, $b_3 = 0.710$, $p < 0.001$), and
(4) Inclusion of concern for reporting the economic substance of the transaction in the analysis reduces the effect of standard precision on aggressive reporting (path IV, $b_4 = 0.740$ and $p = 0.030$ when concern for economic substance is included in the regression versus $b_4 = 2.778$ and $p < 0.001$ when concern for economic substance is *not* included in the regression).

Further, we find that the mediating effect predicted by H3 is significant ($z = 2.35$, $p = 0.009$; see MacKinnon and Dwyer 1993). Thus, this analysis identifies two mechanisms that help explain the

FIGURE 1
Experiment 2: Three-Path Mediation Model and Results



Standard Precision is the treatment variable manipulated at two levels (0 = more precise; 1 = less precise).
Aggressive Financial Reporting is a participant's lease classification decision recorded on a ten-point scale, where 1 = "Definitely classify as an operating lease" and 10 = "Definitely classify as a capital lease." Lower responses are indicative of more aggressive reporting decisions.
Regulator Second-Guessing is a participant's response to the post-experimental question "Relative to other factors, how much was your lease classification decision influenced by your desire to: Avoid possible second-guessing of my decision by external watchdogs such as the Securities and Exchange Commission" (recorded on an 11-point scale where 0 = "Little influence relative to other factors" and 10 = "Very strong influence relative to other factors").
Economic Substance is a participant's response to the post-experimental question "Relative to other factors, how much was your lease classification decision influenced by your desire to: Report the economic substance of the lease in the financial statements" (recorded on an 11-point scale where 0 = "Little influence relative to other factors" and 10 = "Very strong influence relative to other factors").
b_1 through b_4 are the estimated coefficients from the following regression equations:

$$\textit{Regulator Second-Guessing} = \beta_{01} + \beta_1(\textit{Standard Precision}) + \varepsilon_1$$

$$\textit{Economic Substance} = \beta_{02} + \beta_2(\textit{Regulator Second-Guessing}) + \varepsilon_2$$

$$\textit{Aggressive Financial Reporting} = \beta_{03} + \beta_3(\textit{Economic Substance}) + \beta_4(\textit{Standard Precision}) + \varepsilon_3$$

To determine the significance of the mediating relationship (paths I through III), a z-score is calculated using the coefficient values and their standard errors. We calculate the variance of the mediating effect (S_{b1b2b3}) using a first-order Taylor series estimate (multivariate delta method) (MacKinnon and Dwyer 1993): $z = (b_1 b_2 b_3) / S_{b1b2b3} = 2.35$ ($p = .009$).

results we obtained in experiment 1 related to the influence of level of standard precision on aggressive financial reporting (H1).

Other Factors Influencing the Lease Classification Decision

Table 4 presents three other potential influencing factors: concern for second-guessing by the audit committee, concern for second-guessing by the auditor, and desire to present the company's

financial position as favorably as the circumstances allow. Consistent with prior archival work in rules-based settings, audit committee strength appears to have significant influence over participants' decisions when applying a more rules-based standard (means of more precise standard/strong audit committee and more precise standard/weak audit committee = 6.00 and 1.77, respectively, $p < 0.001$).

It is interesting to note, however, that a stronger audit committee appears relatively less influential in our more principles-oriented setting (means of less precise standard/strong audit committee condition and more precise standard/strong audit committee condition = 3.44 and 6.00, respectively, $p < 0.001$). This is consistent with our finding that our principles-based standard eases the burden on audit committees (whether strong or weak) to curtail management's aggressive reporting choices.[20] That is, when applying our more principles-oriented standard, it appears that preparers are concerned primarily about regulator second-guessing. This, in turn, dampens pursuit of aggressive financial reporting and results in reporting choices that are more consistent with audit committee preferences.

Further examination of potential influential factors in the two more precise standard cells suggests that, when the audit committee is weaker, financial statement preparers are more concerned with presenting the company's financial position as favorably as the circumstances will allow (mean = 7.32) than when the audit committee is stronger (mean = 6.41, $p = 0.082$). It also is interesting to note that presenting a favorable financial position has a much smaller influence on the lease classification decision for participants in the less precise standard condition than for participants in the more precise standard condition (means of 3.39 and 6.86, respectively, $p < 0.001$).

V. SUMMARY AND CONCLUSIONS

As a result of recent corporate accounting scandals, the U.S. Congress passed SOX in an effort to reform the U.S. financial reporting system. In turn, this legislation has pushed the SEC and the FASB to consider reforms designed to inhibit aggressive financial reporting. These reforms include a shift toward more principles-based accounting standards and enhancing the role of the audit committee. We provide the first evidence on how these two SOX-related regulatory initiatives jointly impact financial statement preparers' reporting judgments. Specifically, our study reports the results of two experiments investigating the effect of financial reporting standard precision on experienced financial statement preparers' financial reporting judgments, as well as the potential role that audit committee strength plays in mitigating aggressive financial reporting.

Consistent with expectations, our results show that financial statement preparers are less likely to report aggressively when applying a less precise financial reporting standard than when applying a more precise standard. We also find a significant interactive effect of audit committee strength and standard precision on preparers' lease classification decisions. We find that audit committee strength affects aggressive reporting in a rules-based regime by inhibiting financial statement preparers' opportunistic application of a standard. In contrast, audit committee strength has no effect in a more principles-based regime where preparers are more concerned about regulator second-guessing and reporting the economic substance of a transaction regardless of audit committee strength. We also find support for a three-path mediating model in which concern about second-guessing by regulators and the desire to report the economic substance of a transaction sequentially mediate the relationship between standard precision and preparers' aggressive financial reporting. Further, we find significantly less variability among preparers' financial reporting

[20] In contrast to the audit committee (a manipulated variable), we did not expect, nor did we find, significant differences between any of the four cells with respect to concern for second-guessing by the auditor ($p = 0.847$).

decisions when a less precise standard is in place. This suggests that, contrary to the concern of some, the application of more principles-based standards need not result in less inter-firm comparability than more precise standards.

Our findings have implications for both practice and research. With respect to practice, our results provide insight into the principles-versus-rules debate and should be of interest to U.S. policy makers as they continue to contemplate a switch from rules-based to principles-based accounting standards. Our findings suggest that moving toward more principles-based standards (such as the SEC's proposed adoption of IFRS) will not necessarily open the door to greater opportunistic reporting by financial statement preparers; instead, this shift could result in more economically meaningful reporting. Also, our results regarding variability in financial statement preparer responses can help allay concerns regarding inter-firm comparability in a principles-based standards regime. Further, in our setting, we find that switching to a more principles-based approach appears to have a greater dampening effect on aggressive reporting than does strengthening the audit committee. This result, while context-specific, warrants further inquiry by policy makers and researchers. Future research could further explore the relationship between audit committee strength and financial reporting quality under more principles-oriented regimes. Finally, prior research suggests that audit partners believe principles-based standards will reduce their power in resolving auditor-client conflicts (Gibbins et al. 2001). However, our results suggest that the number of such disagreements could diminish, which should help alleviate concerns of auditors' lost leverage.

It is important to note that our experiments focus on a setting in which the rules-based standard has been unable to curb aggressive reporting (i.e., the lease capitalization decision), even given a strong regulatory environment in which there is general concern over aggressive reporting choices. To broaden the generalizability of our findings, future research could explore other settings involving more restrictive rules that more effectively bind financial statement preparers to a conservative treatment. Further research could also explore the effect of differing levels of regulatory scrutiny on principles-based standards. Such research would further our understanding of the roles that standard precision and audit committee strength play in mitigating aggressive financial reporting practices.

REFERENCES

Abbott, L., J. Park, and S. Parker. 2000. The effects of audit committee activity and independence on corporate fraud. *Managerial Finance* 26 (11): 55–67.

———, S. Parker, and G. F. Peters. 2004. Audit committee characteristics and restatements. *Auditing: A Journal of Practice & Theory* 23 (1): 69–87.

Blue Ribbon Committee. 1999. *Report and Recommendations of the Blue Ribbon Committee on Improving the Effectiveness of Corporate Audit Committees*. New York, NY: NYSE and NASD.

Bédard, J., S. M. Chtourou, and L. Courteau. 2004. The effect of audit committee expertise, independence, and activity on aggressive earnings management. *Auditing: A Journal of Practice & Theory* 23 (2): 13–35.

Benston, G. J., M. Bromwich, and A. Wagenhofer. 2006. Principles-versus rules-based accounting standards: The FASB's standard setting strategy. *Abacus* 42 (2): 165–188.

Bockus, K. A., W. D. Northcut, and M. E. Zmijewski. 2003. Accounting and disclosure issues in structured finance. In *Corporate Aftershock: The Public Policy Lessons from the Collapse of Enron and Other Major Corporations*, edited by C. L. Culp and W. A. Niskanen, 193–208. Hoboken, NJ: John Wiley.

Box, G. E. P. 1954. Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification. *Annals of Mathematical Statistics* 25: 290–302.

Calfee, J. E., and R. Craswell. 1984. Some effects of uncertainty on compliance with legal standards. *Virginia Law Review* 70 (5): 965–1003.

Carcello, J. V., C. W. Hollingsworth, and T. L. Neal. 2006. Audit committee financial experts: A closer examination using firm designations. *Accounting Horizons* 20 (4): 351–373.

Chen, J. J., and P. Cheng. 2007. Corporate governance and the harmonization of Chinese accounting practices with IFRS practices. *Corporate Governance* 15 (2): 284–293.

Collier, P., and A. Gregory. 1996. Audit committee effectiveness and the audit fee: *European Accounting Review* 5 (2): 177–198.

Craswell, R., and J. E. Calfee. 1986. Deterrence and uncertain legal standards. *Journal of Law Economics and Organization* 2 (2): 279–303.

Cuccia, A., K. Hackenbrack, and M. Nelson. 1995. The ability of professional standards to mitigate aggressive reporting. *The Accounting Review* 70 (2): 227–248.

DeFond, M. L., R. N. Hann, and X. Hu. 2005. Does the market value financial expertise on audit committees of boards of directors? *Journal of Accounting Research* 43 (2): 153–193.

DeZoort, F. T., D. R. Hermanson, D. S. Archambeault, and S. A. Reed. 2002. Audit committee effectiveness: A synthesis of the empirical audit committee literature. *Journal of Accounting Literature* 21: 38–75.

Financial Accounting Standards Board (FASB). 1976. *Accounting for Leases*. Statement of Financial Accounting Standards No. 13. Stamford, CT: FASB.

———. 2002. *Proposal—Principles-Based Approach to U.S. Standard Setting*. Norwalk, CT: FASB.

Gibbins, M., S. Salterio, and A. Webb. 2001. Evidence about auditor-client management negotiation concerning client's financial reporting. *Journal of Accounting Research* 39 (3): 535–563.

———, S. A. McCracken, and S. E. Salterio. 2007. The chief financial officer's perspective on auditor client negotiations. *Contemporary Accounting Research* 24 (2): 387–422.

Graham, J. R., and C. R. Harvey. 2001. The theory and practice of corporate finance: Evidence from the field. *Journal of Financial Economics* 60: 187–243.

Hackenbrack, K., and M. W. Nelson. 1996. Auditors' incentives and their application of financial accounting standards. *The Accounting Review* 71 (1): 43–59.

Hoffman, V. B., and J. M. Patton. 2002. How are loss contingency accruals affected by alternative financial reporting criteria and incentives? *Journal of Accounting and Public Policy* 21: 151–167.

Huron Consulting Group (Huron). 2006. *2006 Audit Committee Research Report*. Chicago, IL: Huron Consulting Services LLC.

Imhoff, E., and J. Thomas. 1988. Economic consequences of accounting standards: The lease disclosure rule change. *Journal of Accounting and Economics* 10 (4): 277–310.

International Accounting Standards Committee (IASC). 1997. *Leases*. International Accounting Standard 17. London, U.K.: IASC.

Jamal, K., and H. T. Tan. 2010. Joint effects of principles-based versus rules-based standards and auditor type in constraining financial managers' aggressive reporting. *The Accounting Review* 85 (4): 1325–1346.

Kadous, K., S. Kennedy, and M. Peecher. 2003. The effect of quality assessment and directional goal commitment on auditors' acceptance of client-preferred accounting methods. *The Accounting Review* 78 (3): 759–778.

Klein, A. 2002. Economics determinants of audit committee independence. *The Accounting Review* 77 (2): 435–452.

Lander, G. P. 2004. *What is Sarbanes-Oxley?* Boston, MA: McGraw-Hill.

Libby, R., R. Bloomfield, and M. W. Nelson. 2002. Experimental research in financial accounting. *Accounting, Organizations and Society* 27 (8): 775–810.

Lindman, H. R. 1974. *ANOVA in Complex Experimental Designs*. San Francisco, CA: WH Freeman.

MacKinnon, D., and J. Dwyer. 1993. Estimating mediated effects in prevention studies. *Evaluation Review* 17 (April): 144–158.

Maines, L. A., E. Bartov, P. Fairfield, D. E. Hirst, T. E. Iannaconi, R. Mallett; C. M. Schrand, D. J. Skinner, and L. Vincent. 2003. Evaluating concepts-based vs. rules-based approaches to standard setting. *Accounting Horizons* 17 (1): 73–89.

——— 2007. Spotlight on principles-based financial reporting. *Business Horizons* 50 (5): 359–364.

Nelson, M. W., J. A. Elliott, and R. L. Tarpley. 2002. Evidence from auditors about managers' and auditor's earnings management decisions. *The Accounting Review* 77 (Supplement): 175–202.

——— 2003. Behavioral evidence on the effects of principles- and rules-based standards. *Accounting Horizons* 17 (1): 91–104.

Ng, T. B., and H.-T. Tan. 2003. Effects of authoritative guidance availability and audit committee effectiveness on auditors' judgments in an auditor-client negotiation context. *The Accounting Review* 78 (3): 801–818.

Peasnell, K. V., P. Pope, and S. Young. 2000. Accrual management to meet earnings targets: UK evidence pre- and post-Cadbury. *The British Accounting Review* 32 (4): 415–445.

Psaros, J., and K. T. Trotman. 2004. The impact of the type of accounting standards on preparers' judgments. *Abacus* 40 (1): 76–93.

Pulliam, S. 1988. Beating FAS 13. *Corporate Finance* (December): 31.

Sanchez, M. H., C. P. Agoglia, and R. C. Hatfield. 2007. The effect of auditors' use of a reciprocity-based strategy on auditor-client negotiations. *The Accounting Review* 82 (1): 241–263.

Schipper, K. 2003. Principles-based accounting standards. *Accounting Horizons* 17 (1): 61–72.

Securities and Exchange Commission (SEC). 2003. *Study Pursuant to Section 108(d) of the Sarbanes Oxley Act of 2002 on the Adoption by the United States Financial Reporting System of a Principles-Based Accounting System*. Washington, D.C.: SEC.

———. 2008. *Roadmap for the Potential Use of Financial Statements Prepared in Accordance With International Financial Reporting Standards by U.S. Issuers*. Washington, D.C.: SEC.

———. 2010. *Commission Statement in Support of Convergence and Global Accounting Standards*. Washington, D.C.: SEC.

Shortridge, R. T., and M. Myring. 2004. Defining principles-based accounting standards. *The CPA Journal Online*, available at: www.nysscpa.org/cpajournal/2004/804/essentials/p34.htm.

Stewart, J., and L. Munro. 2007. The impact of audit committee existence and audit committee meeting frequency on the external audit: Perceptions of Australian auditors. *International Journal of Auditing* 11 (1): 51–69.

Trompeter, G. 1994. The effect of partner compensation schemes and generally accepted accounting principles on audit partner judgment. *Auditing: A Journal of Practice & Theory* 13 (2): 56–68.

U.S. House of Representatives. 2002. *The Sarbanes-Oxley Act of 2002. Public Law No. 107–204 [H.R. 3763]*. Washington, D.C.: Government Printing Office.

Weil, R. 2002. *Fundamental Causes of the Accounting Debacle at Enron: Show Me Where It Says I Can't*. Washington, D.C.: The Committee on Energy and Commerce, U.S. House of Representatives.

Xie, B., W. N. Davidson, and P. J. DaDalt. 2003. Earnings management and corporate governance: The role of the board and the audit committee. *Journal of Corporate Finance* 9: 295–316.

THE ACCOUNTING REVIEW
Vol. 86, No. 3
2011
pp. 769–804

American Accounting Association
DOI: 10.2308/accr.00000033

Effects of SFAS 133 on the Risk Relevance of Accounting Measures of Banks' Derivative Exposures

Anwer S. Ahmed
Texas A&M University

Emre Kilic
Gerald J. Lobo
University of Houston

ABSTRACT: We provide evidence on the effects of SFAS 133 on the risk relevance of accounting measures of bank derivative exposures to bond markets. First, we find that interest rate derivatives classified as hedging are more *negatively* associated with fixed-rate bond spreads after SFAS 133. We also find that hedging derivatives offset non-trading positions to a greater extent after SFAS 133. Second, for the largest 25 banks, we find that interest and foreign exchange rate trading derivatives are more *negatively* associated with fixed-rate bond spreads after SFAS 133, consistent with more economic hedges being classified as trading after SFAS 133. For these banks, trading derivative exposures offset non-derivative trading exposures to a greater extent after SFAS 133. Our results suggest that, contrary to critics' claims, SFAS 133 has increased the risk relevance of accounting measures of derivative exposures to bond investors and benefited banks in terms of reducing their cost of capital.

Keywords: *risk relevance; derivatives; SFAS 133; bond spread.*

JEL Classifications: *G12; G14; G21; G32; M41.*

I. INTRODUCTION

We provide evidence on the effects of the Statement of Financial Accounting Standards (SFAS) 133, *Accounting for Derivative Instruments and Hedging Activities* (FASB 1998), on the risk relevance of accounting measures of bank derivative

We gratefully acknowledge comments from Mary Lea McAnally, David Harris, Tom Lys, Douglas Stevens, Martin Wu, Yildiray Yildirim, and workshop participants at the University of Illinois–Urbana Champaign, Georgia State University, Syracuse University, Florida State University, and Virginia Commonwealth University.

Editor's note: Accepted by Steven Kachelmeier.

Submitted: July 2009
Accepted: September 2010
Published Online: April 2011

exposures.[1] We focus on banks because they extensively use derivatives in both hedging and trading activities. We assess risk relevance by examining the extent to which bank fixed-rate bond spreads (hereafter, bond spreads) are associated with measures of exposures from derivatives classified as hedging (hereafter, hedging derivatives) and derivatives classified as trading (hereafter, trading derivatives). We also provide corroborating evidence on how SFAS 133 affected the risk-reducing (or risk-enhancing) role of derivatives by studying the association between changes in fair values of derivatives and changes in fair values of non-derivative positions.

These research questions are important for at least three reasons. First, even though accounting guidance for derivatives prior to SFAS 133 was inconsistent, incomplete, and lacking in transparency, there was considerable controversy over the potential benefits and costs of SFAS 133, especially for banks (see Section II for more discussion of the accounting guidance before and after SFAS 133). In particular, proponents argued that, because SFAS 133 provides greater uniformity in accounting for derivatives and standardizes the criteria for qualifying for hedge accounting, it improves the transparency and reliability of derivatives reporting by banks as well as their monitoring of risk-management strategies involving derivatives, thereby reducing their cost of capital.[2] However, critics argued that SFAS 133 makes it more difficult to obtain hedge accounting treatment and exacerbates the difference between accounting and economic hedges.[3] Thus, to the extent banks record economic hedges as if they are speculative and investors penalize banks for these positions (Morgan and Stiroh 2001), SFAS 133 would increase banks' cost of capital. Similar views were expressed by the Financial Executives Institute (1997, 2), which stated, "[SFAS 133] will send a highly misleading message to readers of financial statements, unfavorably affecting the cost of capital." Because one of the precepts the Financial Accounting Standards Board (FASB) follows when issuing a standard is that "the costs imposed to meet the standard should be justified in relation to the overall benefits of the resulting information" (SFAS 133, ¶ 232), documenting evidence on the effects of SFAS 133 is necessary for a complete cost-benefit analysis of this standard.[4]

Second, prior accounting studies focus mostly on the value relevance of bank derivative disclosures (Barth et al. 1996; Eccher et al. 1996; Nelson 1996; Venkatachalam 1996; Ahmed et al. 2006), risk relevance of bank income measures (Beaver et al. 1970; Hodder et al. 2006), and risk relevance of disclosures to equity investors (McAnally 1996; Schrand 1997). To our knowledge, the risk relevance of accounting measures of derivative exposures to bond investors has not been examined previously. This is an important gap in the literature because (1) bonds are a much larger source of capital than stocks,[5] and (2) there are important structural differences between bond and

---

[1] SFAS 138 (FASB 2000), Accounting for Certain Derivative Instruments and Certain Hedging Activities—An amendment of SFAS 133, and SFAS 149 (FASB 2003), Amendment of SFAS 133 on Derivative Instruments and Hedging Activities, which were issued later, address certain issues related to the implementation of SFAS 133, but do not change the basic model of SFAS 133.

[2] For example, the Investor Protection Department of the Consumer Federation of America (1997, 1) states that: "[accounting standards for derivatives and hedging activities *before* SFAS 133] leave investors in the dark about billions of dollars of transactions that the companies they invest in engage in each year. [SFAS 133] will aid investors by increasing the visibility, comparability, and understandability of the risks and benefits associated with these often complex financial instruments."

[3] For example, Canadian Bankers Association (1997, 2) states that "The hedging criteria in [SFAS 133] are overly restrictive and disallow hedge accounting in situations where economic exposures are currently being effectively hedged."

[4] The responses to the Exposure Draft of SFAS 133 indicate that banks generally opposed SFAS 133 from a cost-benefit standpoint. For example, First Chicago NBD Corporation (1997, 13–14) states: "The standard cannot be justified from a cost-benefit standpoint … [Proposed changes] are not critical enough to compensate for the cost of implementing the standard."

[5] Flow of Funds data published by the Board of Governors of the Federal Reserve System indicate that the size of the debt market at the end of 2005 was about twice that of the equity market.

stock markets. For example, bond markets are dominated by more sophisticated participants than are stock markets (Bhojraj and Swaminathan 2009), but are less liquid than stock markets. This study jointly investigates the risk relevance of accounting measures of derivative exposures in bond markets and how SFAS 133 altered that relevance.

Third, studying the effects of SFAS 133 on bond spreads is a more direct and cleaner way of studying the effects of SFAS 133 on the risk relevance of derivatives reporting because the primary driver of bond prices is risk, whereas stock prices are driven by changes in expected future cash flows as well as changes in risk (or discount rates).

If derivatives are used for risk reduction, then derivative exposures would be negatively associated with bond spreads. We hypothesize that derivatives classified as hedging are more negatively associated with bond spreads after SFAS 133 because SFAS 133 uses more stringent criteria for classifying derivatives as hedging derivatives and requires more detailed documentation of the hedging relationship than prior standards, thus increasing the likelihood that hedging derivatives consist of highly effective derivative hedges. On the other hand, trading derivatives will be positively (negatively) associated with bond spreads if they increase (reduce) net exposures. However, because the relative weight of derivatives used for economic hedges within the trading derivatives category is likely to increase after SFAS 133, we also hypothesize that trading derivatives will be less positively (or more negatively) associated with bond spreads after SFAS 133 if bondholders correctly perceive the increase in the relative weight of economic hedges within the trading derivatives category. We expand on these arguments in Section II.

We examine the association between bank bond spreads and accounting measures of derivative exposures before and after SFAS 133 while controlling for bond and bank characteristics identified in prior studies. Our tests are based on a sample of subordinated bank bonds with U.S. dollar denominations. We find that, after SFAS 133, exposures from interest rate hedging derivatives are more negatively associated with bond spreads. In terms of economic magnitudes, before SFAS 133, the tax-adjusted bond spread is approximately 19.5 basis points lower for a bank with the mean level of exposure from hedging derivatives relative to a bank with no such exposure, whereas it is 30.8 basis points lower after the adoption of SFAS 133. Furthermore, we find that exposures from interest and foreign exchange rate trading derivatives are also more negatively associated with bond spreads after SFAS 133 for the largest banks, which are extensively involved in derivatives trading. For these banks, the mean level of exposure from trading derivatives is associated with a 5.8 basis point lower spread before, and a 9.9 basis point lower spread after SFAS 133, compared to banks without such exposure.

To provide corroborating evidence on the risk-reducing role of derivatives, we investigate how SFAS 133 altered the relation between fair values of derivative and non-derivative exposures. Specifically, we hypothesize that, after SFAS 133, (1) hedging derivatives offset exposures from non-trading positions to a greater extent, and (2) trading derivatives offset exposures from non-derivative trading positions to a greater extent. Consistent with these hypotheses, we find that fair value changes in hedging derivatives and trading derivatives are more negatively related to fair value changes in non-trading positions and non-derivative trading positions, respectively, after SFAS 133. These results suggest that hedging derivatives offset exposures from non-trading positions, whereas trading derivatives offset exposures from non-derivative trading positions.

We contribute to the prior literature in two important ways. First, SFAS 133 has been one of the most controversial standards ever issued by the FASB. Contrary to critics' claims, we find that accounting measures of derivative exposures are more negatively associated with bond spreads after SFAS 133 than before SFAS 133, indicating that SFAS 133 has improved the risk relevance of derivatives reporting, heightened investors' confidence in the stated purposes of holding derivatives, and benefited banks in terms of reducing their cost of capital.

Second, prior work does not provide evidence on how bond investors price derivative exposures. We document that bond investors price both hedging and trading derivatives in a manner consistent with their use in offsetting non-derivative inherent risk exposures, and that these offsetting effects are more pronounced after SFAS 133. Our findings contrast with the results of prior studies (such as Barth et al. 1996; Eccher et al. 1996; Nelson 1996; Ahmed et al. 2006) that find that equity investors do not price derivative disclosures prior to SFAS 133, but are consistent with the greater sophistication of bond market investors.

Section II presents the background and hypotheses development. Section III describes the research design. Section IV presents the evidence and Section V the conclusion.

II. BACKGROUND AND HYPOTHESES DEVELOPMENT

Hedge Accounting for Derivatives before and after SFAS 133

Prior to SFAS 133, accounting for derivatives was primarily guided by SFAS 52 (FASB 1981), *Foreign Currency Translation*, and SFAS 80 (FASB 1984), *Accounting for Futures Contracts*, which addressed foreign currency transactions and exchange-traded futures contracts, respectively, and the Emerging Issues Task Force (EITF) Issue No. 84-36. There were several important inconsistencies in the requirements for hedge accounting treatment between SFAS 52 and SFAS 80.[6] First, SFAS 52 required transaction-level risk reduction for derivatives to be classified as hedging, whereas SFAS 80 required enterprise-level risk reduction. Second, while SFAS 52 did not refer to correlation between changes in the values of hedging derivatives and the hedged items, SFAS 80 frequently emphasized a high correlation requirement for an effective hedge (e.g., ¶ 4 (b) and 11). Consequently, the Securities and Exchange Commission (SEC), under SFAS 80, interpreted a hedge to be effective if the cumulative change in the value of the hedging instrument was between 80 and 125 percent of the cumulative change in the value of the hedged item, but did not articulate a specific threshold for assessing hedge effectiveness under SFAS 52 or EITF 84-36 (Swad 1995; Carpenter 1996; Anson 1999; Finnerty and Grant 2002). Third, SFAS 52 did not require ongoing assessment of hedge effectiveness, whereas SFAS 80 did. Finally, SFAS 52 did not allow hedging of anticipated transactions, whereas SFAS 80 allowed such hedging.

In addition to these inconsistencies, there were also gaps in the accounting guidance. Neither SFAS 52 nor SFAS 80 addressed widely used derivatives such as interest rate swaps and options (Beier 1993; Montesi and Lucas 1996). Hedge accounting for swaps and options was applied by analogy to SFAS 52 and SFAS 80 (Gastineau et al. 2001). However, depending upon the sources of the analogy, the classification of and accounting for derivatives were *ad hoc* and often inconsistent across reporting entities (SFAS 133, ¶ 237). Some derivatives were accounted for as hedges regardless of whether they were part of a hedging strategy (SFAS 133, ¶ 235) or whether they were effective hedges.

The EITF attempted to address the accounting for interest rate swaps with EITF 84-36. However, in contrast to the guidance in SFAS 80, enterprise-level interest rate risk reduction was not required by EITF 84-36. Hence, prior to SFAS 133, the inconsistencies between the effective standards regarding what constituted a hedge, coupled with the failure of those standards to cover widely used derivatives contracts, made disclosure of and accounting for derivatives among the least standardized areas of financial reporting.

SFAS 133 was issued to provide uniformity in accounting for derivatives and standardize hedge accounting criteria.[7] Regarding the former, the standard requires the recognition of all

---

[6] Rane (1992), Montesi and Lucas (1996), Anson (1999), Gastineau et al. (2001) and the FASB (in SFAS 133) present detailed discussions of the inconsistencies between SFAS 52 and SFAS 80.

[7] In order to increase transparency in derivatives activities, the FASB adopted SFAS 119 (FASB 1994), *Disclosure about*

derivatives at fair value and recognition of changes in the fair value of derivatives either in net income or in other comprehensive income. As for the latter, SFAS 133 standardized the criteria to qualify for hedge accounting and made what constitutes a hedge more precise. Most relevant to banks, under SFAS 133, derivatives that hedge certain types of risks (i.e., the risk of changes in the value or cash flows) from specific exposures (i.e., individual assets or liabilities) are classified as hedging and qualify for hedge accounting.[8] In addition, SFAS 133 requires formal and unambiguous documentation of the hedging relationship between the derivative and the specific exposure.[9] Given that documentation of the relation between the hedged and the hedging exposures at the enterprise level is difficult, a shift in focus from enterprise-level risk reduction to specific-exposure risk reduction made the relation between the hedging and the hedged exposures easier to document and therefore more identifiable after SFAS 133.

While SFAS 133 does not provide a specific hedge effectiveness threshold, similar to the SEC's effectiveness interpretation under SFAS 80, practitioners interpret high effectiveness to mean that cumulative changes in the value of the hedging instrument should be between 80 and 125 percent of the inverse cumulative changes in the value or the cash flows of the hedged item. Accordingly, most macro hedging derivatives do not qualify for hedge accounting under SFAS 133 as there is no objective method of gauging the effectiveness of the hedging derivative without linkage to a single, identifiable asset or liability (SFAS 133, ¶ 447).

Hypothesis Development

Corporate risk management theory posits that hedging reduces costs associated with external financing (Mayers and Smith 1982; Smith and Stulz 1985; Dadalt et al. 2002). Banks hold hedging derivatives to offset exposures mainly to macroeconomic risk factors, particularly interest and foreign exchange rates. Therefore, such derivatives reduce a bank's systematic risk by making its cash flows less dependent on macroeconomic factors (Froot et al. 1993; Gay and Nam 1998), and reduce a bank's default risk by shielding its capital from losses due to unanticipated movements in market risk factors. Given the finding of Elton et al. (2001) that bond spreads consist mainly of systematic and default risk premia after tax premium is removed, derivative hedges likely result in lower bond spreads to the extent they are effective.

As discussed earlier, prior to SFAS 133, many derivatives were accounted for under EITF 84-36 or by analogy to SFAS 52, both of which conflicted with SFAS 80 in terms of risk reduction and did not contain an explicit hedge effectiveness criterion. Derivatives classified as "hedging" after SFAS 133 likely meet a higher hedge effectiveness threshold "on average" than do derivatives classified as hedging before SFAS 133 because SFAS 133 (1) links hedging derivatives with specific identifiable exposures, (2) standardizes the hedge effectiveness criteria, and (3) requires explicit documentation and ongoing assessment of the hedging relationship. The more stringent qualification and documentation rules imposed by SFAS 133 are likely to increase investors' confidence in the stated purpose of holding hedging derivatives and improve the monitoring of hedging activities, resulting in less uncertainty in the financial statements and lower bank bond spreads. This leads to our first hypothesis:

Derivative Financial Instruments and Fair Value of Financial Instruments. SFAS 119, which was superseded by SFAS 133, regulated only the derivative-related disclosures, but neither changed the existing accounting for derivatives nor the criteria for qualifying as a hedging derivative. Instead, it relied on the classification criteria provided by SFAS 52 and SFAS 80.

8 SFAS 133 also allows hedging of off-balance sheet firm commitments and forecasted transactions.

9 Documentation under SFAS 133 should include identification of (1) the hedging derivative, (2) the hedged asset or liability, (3) the nature of the risk being hedged, and (4) how the effectiveness of the hedge is assessed.

H1: Bank bond spreads are more negatively related to derivatives classified as hedging after SFAS 133 than before SFAS 133.

Bank trading derivatives consist of customer-related positions, proprietary positions, and economic hedges. Banks typically take offsetting positions in customer-related derivatives trading. By matching customer-related derivatives in terms of market risk exposure, a bank can generate fee income without assuming any market risk. Although not as common, some banks also take non-offset positions in customer-related derivatives trading (Puwalski 2003). Hence, whether customer-related derivatives increase or reduce overall market risk depends on whether the residual market risk, if any, from these positions offsets or adds to other risk exposures. Unlike customer-related derivatives, proprietary trading derivatives are likely to increase exposure because they are held for speculation. Lastly, some derivatives classified as trading are in fact economic hedges of other exposures.[10] Therefore, bank derivatives classified as trading include risk-reducing positions (i.e., economic hedges), risk-enhancing positions (i.e., speculative positions), and positions with ambiguous risk implications (i.e., customer-related positions). However, the relative weights of these positions are not disclosed under SFAS 119 (FASB 1994) or under SFAS 133. Thus, the relation between bond spreads and trading derivatives could be negative or positive, depending upon whether the net trading derivatives positions reduce or increase overall risk. However, the relative importance of economic hedges within trading derivatives is likely to increase after SFAS 133 for at least two reasons. First, as discussed earlier, SFAS 133 prohibits the reporting of derivatives used in macro hedging as hedging derivatives, and requires them to be classified as trading derivatives. Second, some derivatives used by banks in *de facto* hedging transactions are classified as trading because they are no longer defined as hedging derivatives, or are not sufficiently effective under SFAS 133. Such derivatives include hedges of interest rate risk in held-to-maturity securities (SFAS 133, ¶ 21 (d) and ¶ 29 (e)), long-term core deposits (SFAS 133, ¶ 317 and ¶ 437), and prepayment risks of financial assets or liabilities (SFAS 133, ¶ 29 (h)).

Banks commonly macro-hedge gap positions or value-at-risk. Therefore, they opposed the disqualification of macro hedges from hedge accounting in their response letters to the Exposure Draft for SFAS 133. The general view of the banking industry was that reporting some legitimate economic hedges as trading, as required by SFAS 133, would generate potentially misleading and unfavorable results when applied to banks (America's Community Bankers 1997). While these economic hedges are typically less effective hedges than the derivatives that qualify for hedge accounting under SFAS 133, they are more effective hedges than derivatives classified as trading prior to SFAS 133. Therefore, if creditors price trading derivatives in a manner consistent with these derivatives including more economic hedges after SFAS 133, a more negative (or less positive) association between bank bond spreads and trading derivatives will result. Thus, we test the following hypothesis:

H2: Bank bond spreads are more negatively (or less positively) related to derivatives classified as trading after SFAS 133 than before SFAS 133.

Derivatives that hedge exposures from trading positions are required to be classified as trading (SFAS 119, ¶ 47), whereas derivatives that hedge exposures from non-trading assets and liabilities, such as loans and deposits, are to be classified as hedging. One of the presumptions underlying H1 is that derivatives classified as hedging after SFAS 133 are, on average, more effective hedges than

---

[10] For example, in its 10-K for 1999, Compass Bancshares, Inc. states: "The primary purposes for using interest rate derivatives in the trading account are to facilitate customer transactions and to *help protect cash market positions in the trading account against interest rate movement.* [Some options in trading account] assist in *protecting the fair value of the trading account* against rising short-term interest rates while maintaining limited risk to declining rates." (emphasis added)

derivatives classified as hedging before SFAS 133 because SFAS 133 restricts derivatives classified as hedging to highly effective derivative hedges of specific exposures. Therefore, we expect derivatives classified as hedging to offset the corresponding hedged items to a greater extent after SFAS 133, which leads to the following hypothesis:

H3: The relation between change in fair value of derivatives classified as hedging and change in fair value of net non-trading assets is more negative after SFAS 133 than before SFAS 133.

Unlike fair values of hedging derivatives, fair values of trading derivatives need not covary with other existing exposures in any specific direction to be classified as trading under either SFAS 119 or SFAS 133. Therefore, whether trading derivatives offset or add to the exposures from non-derivative trading or non-trading positions is an empirical question. Liu et al. (2004) do not find significant cross-hedging between trading and non-trading exposures. On the other hand, trading derivatives include derivatives that are economic hedges of other trading positions. SFAS 133 likely increases the average weight of such economic hedges in bank derivative portfolios classified as trading for two reasons. First, as mentioned earlier, some derivatives that are classified as hedging prior to SFAS 133 are classified as trading after SFAS 133. Because recognizing the hedging derivative at fair value, but the hedged position at cost would create income volatility, SFAS 133 allows banks to transfer such hedged items into the trading category along with derivatives hedging these positions (SFAS 133, ¶ 54). This likely increases the weight of derivative and non-derivative positions with offsetting exposures in the trading portfolio.[11] Second, banks could choose to avoid designating some derivatives as hedging and instead classify them as trading after SFAS 133 because the costs associated with the more onerous documentation and more complex hedge accounting requirements under SFAS 133 may outweigh the benefits.[12] This reasoning leads to the following hypothesis:

H4: The relation between change in fair value of derivatives classified as trading and change in fair value of net non-derivative trading assets is more negative (or less positive) after SFAS 133 than before SFAS 133.

III. RESEARCH DESIGN AND EMPIRICAL MODELS

Model for Testing Hypotheses 1 and 2

Hypotheses 1 and 2 posit that the relation between accounting measures of derivative exposures and bond spreads will differ before and after SFAS 133. We test these hypotheses with a model that relates accounting measures of derivative exposures to bond spreads, controlling for bond characteristics (Hancock and Kwast 2001),[13] bank characteristics (Flannery and Sorescu 1996; Morgan and Stiroh 2001; Hancock and Kwast 2001; Jagtiani et al. 2002), and off-balance sheet instruments other than derivatives that are potentially used as economic hedges. We estimate the model separately for the periods before and after SFAS 133. The model is specified as follows:

$$
\begin{aligned}
r_{it} = {} & \beta_0 + \beta_1 Big_{it}^{25} + \beta_2 IrHdg_{it} + \beta_3 Big_{it}^{25} * IrHdg_{it} + \beta_4 DeltaIrHdg_{it} + \beta_5 Big_{it}^{25} * DeltaIrHdg_{it} \\
& + \beta_6 FxHdg_{it} + \beta_7 Big_{it}^{25} * FxHdg_{it} + \beta_8 DeltaFxHdg_{it} + \beta_9 Big_{it}^{25} * DeltaFxHdg_{it} \\
& + \beta_{10} IrTrd_{it} + \beta_{11} Big_{it}^{25} * IrTrd_{it} + \beta_{12} DeltaIrTrd_{it} + \beta_{13} Big_{it}^{25} * DeltaIrTrd_{it} + \beta_{14} FxTrd_{it}
\end{aligned}
$$

---

[11] Reclassification of some securities is allowed under SFAS 133 to mitigate the likely effect of the standard on earnings volatility. Recognizing both the derivatives and the hedged items at fair value after SFAS 133 allows the offsetting gains and losses from the hedging and the hedged positions to be recognized in income concurrently.

[12] In line with this view, a survey by Bank of America (2006) revealed that a small but growing number of banks opted not to apply hedge accounting to certain derivative hedges because the cost and effort of effectiveness testing and documentation required under SFAS 133 are too high.

[13] We restrict the sample to subordinated bank bonds with U.S. dollar denominations that are all non-callable, non-putable, non-sinking fund, and non-convertible. Therefore, bonds in our sample vary in maturity and amount only.

$$+ \beta_{15}Big_{it}^{25} * FxTrd_{it} + \beta_{16}DeltaFxTrd_{it} + \beta_{17}Big_{it}^{25} * DeltaFxTrd_{it} + \beta_{18}Maturity_{it}$$
$$+ \beta_{19}Amount_{it} + \beta_{20}Size_{it} + \beta_{21}Insured_{it} + \beta_{22}Estate_{it} + \beta_{23}Npl_{it} + \beta_{24}Gap_{it}$$
$$+ \beta_{25}NonDerTrd_{it} + \beta_{26}Cap_{it} + \beta_{27}Roa_{it} + \beta_{28}CrRisk_{it} + \beta_{29}LoanComm_{it}$$
$$+ \beta_{30}Guarantee_{it} + \beta_{31}CrDer_{it} + \beta_{32}Strip_{it} + \beta_{33}Iml_{it} + \varepsilon_{it}, \quad (1)$$

where for bank i at time t:

r_{it} = tax-adjusted spread;
Big_{it}^{25} = 1 if the bank is one of the largest 25 banks in the U.S. with respect to asset size, and 0 otherwise;
$IrHdg_{it}$ = absolute value of net long notional amount of interest rate derivatives classified as hedging scaled by total assets;
$FxHdg_{it}$ = absolute value of net long notional amount of foreign exchange derivatives classified as hedging scaled by total assets;
$DeltaIrHdg_{it}$ = absolute value of the annual change in the fair value of interest rate derivatives classified as hedging divided by the annual change in the interest rates scaled by total assets;
$DeltaFxHdg_{it}$ = absolute value of the annual change in the fair value of foreign exchange derivatives classified as hedging divided by the annual change in the foreign exchange rates scaled by total assets;
$IrTrd_{it}$ = absolute value of net long notional amount of interest rate derivatives classified as trading scaled by total assets;
$FxTrd_{it}$ = absolute value of net long notional amount of foreign exchange derivatives classified as trading scaled by total assets;
$DeltaIrTrd_{it}$ = absolute value of the annual change in the fair value of interest rate derivatives classified as trading divided by the annual change in the interest rates scaled by total assets;
$DeltaFxTrd_{it}$ = absolute value of the annual change in the fair value of foreign exchange derivatives classified as trading divided by the annual change in the foreign exchange rates scaled by total assets;
$Maturity_{it}$ = natural logarithm of the bond's years to maturity;
$Amount_{it}$ = natural logarithm of the size of bond issue in million dollars;
$Size_{it}$ = natural logarithm of bank's total assets;
$Insured_{it}$ = ratio of total insured deposits to total deposits;
$Estate_{it}$ = real estate acquired in satisfaction of debt previously contracted scaled by total assets;
Npl_{it} = sum of nonaccrual loans and accruing loans past due 90 days or more scaled by total assets;
Gap_{it} = absolute value of the difference between bank assets and liabilities maturing or repricing within one year scaled by total assets;
Roa_{it} = annual net income scaled by total assets;
Cap_{it} = ratio of book value of equity to risk-weighted assets;
$NonDerTrd_{it}$ = fair value of non-derivative trading assets scaled by total assets;
$CrRisk_{it}$ = bank's credit losses incurred on off-balance sheet derivative contracts scaled by total assets;
$LoanComm_{it}$ = amount of loans for which official promises to lend are conveyed scaled by total assets;

$Guarantee_{it}$ = on- and off-balance sheet guarantees and insurance contracts that back financial obligations scaled by total assets;
$CrDer_{it}$ = notional amount of credit derivatives on which the bank is the beneficiary scaled by total assets;
$Strip_{it}$ = fair value of principal-only strips scaled by total assets; and
Iml_{it} = Inverse Mills ratio obtained from the first-stage multinomial logit model.

We provide the list of data sources and discuss the rationale for including each of the control variables in Appendix A.

Interest rate swaps and options constitute the majority of bank derivative portfolios. Ryan (2007) indicates that nonlinearity of option payoff structure makes the assessment of options' effectiveness as hedging instruments difficult when the hedged item has a linear payoff structure. He also indicates that bank market risk disclosures, which are typically linear characterizations of exposures, cannot sufficiently reflect nonlinear exposures. Also, accounting for different types of derivatives under different pronouncements (e.g., forward under SFAS 52, interest rate swaps under EITF 84-36) resulted in differences in hedge accounting for different types of derivatives prior to SFAS 133. Therefore, the effect of SFAS 133 could vary based on the type of derivative. In additional sensitivity analysis, we examine whether a particular type of interest rate derivative drives our results. We re-estimate the model presented in Equation (1) only for interest rate derivatives while allowing the coefficients on (linear) swaps and (non-linear) options to vary.

Since most of the derivatives activity in the U.S. is dominated by the largest 25 banks, we allow the coefficients on the derivative exposure variables to vary across the largest 25 banks and the remaining banks in order to control for the potential difference in risk relevance of information on derivative exposures for these two groups of banks. As hypothesized in H1, we expect more negative coefficients on the accounting measures of exposures from derivatives classified as hedging (i.e., $\beta_2 + \beta_3$, $\beta_4 + \beta_5$, $\beta_6 + \beta_7$, and $\beta_8 + \beta_9$ for the largest 25 and β_2, β_4, β_6, and β_8 for the remaining banks) after the adoption of SFAS 133. Also, as discussed before, whether trading derivatives are positively or negatively related to bond spreads is an empirical question. However, as hypothesized in H2, we expect the coefficients on the accounting measures of exposures from derivatives classified as trading (i.e., $\beta_{10} + \beta_{11}$, $\beta_{12} + \beta_{13}$, $\beta_{14} + \beta_{15}$, and $\beta_{16} + \beta_{17}$ for the largest 25, and β_{10}, β_{12}, β_{14}, and β_{16} for the remaining banks) to be more negative (or less positive) after SFAS 133.

Given our research questions, our study naturally focuses on the subsample of banks that use derivatives. To the extent that the characteristics of derivative-user banks systematically differ from those of non-derivative-user banks, an empirical model that ignores the non-randomness of this subsample likely suffers from selection bias. We control for the potential effects of selection bias by using a standard extension of the Heckman two-stage procedure generalized to polychotomous choices developed by Lee (1983), where the bond-pricing model presented in Equation (1) constitutes the second stage. In the first stage, we classify banks into three categories: banks without derivatives (*Nonuser*), banks with only hedging derivatives (*Hedger*), and banks with both hedging and trading derivatives (*Trader*). Because *Trader* banks are typically those that engage in derivatives activities extensively, the last two categories jointly control for the extent of derivative use and the type of activity (i.e., hedging versus trading). We include the Inverse Mills ratios (*Iml*) obtained from the multinomial logit model in the bond-pricing model.

Finance theory posits that banks with smaller asset portfolio size (*Size*), lower leverage (*Lev*), lower interest rate risk (*Gap*), lower non-performing loans (*Npl*), less convex tax functions (*Convexity*), higher net interest margin (*Nim*), and stronger managerial incentives to take risk (*Incentive*) are less likely to use derivatives for hedging purposes. Prior finance studies also suggest that the likelihood of banks to use derivatives for trading purposes increases with *Size*, liquidity

(*Liquidity*), the size of the commercial and industrial loan portfolio relative to total assets (*ComIndLoan*), and decreases with *Lev* and *Nim*.[14] We specify *Hedger* as the base category in the following first-stage multinomial logit model:

$$D_{it}(Nonuser/Hedger) = \alpha_0 + \alpha_1 Big^{25}_{it} + \alpha_2 Size_{it} + \alpha_3 Lev_{it} + \alpha_4 Gap_{it} + \alpha_5 Nim_{it} + \alpha_6 Npl_{it} + \alpha_7 Liquidity_{it} + \alpha_8 Incentive_{it} + \alpha_9 Convexity_{it} + \alpha_{10} ComIndLoan_{it} + \varepsilon_{it} \quad (2a)$$

$$D_{it}(Trader/Hedger) = \theta_0 + \theta_1 Big^{3}_{it} + \theta_2 Big^{4-25}_{it} + \theta_3 Size_{it} + \theta_4 Lev_{it} + \theta_5 Gap_{it} + \theta_6 Nim_{it} + \theta_7 Npl_{it} + \theta_8 Liquidity_{it} + \theta_9 Incentive_{it} + \theta_{10} Convexity_{it} + \theta_{11} ComIndLoan_{it} + \varepsilon_{it}, \quad (2b)$$

where, for bank i at time t, D_{it} equals 0 if bank i is a *Hedger* in both Equations (2a) and (2b), equals 1 in Equation (2a) if bank i is a *Nonuser*, and equals 1 in Equation (2b) if bank i is a *Trader*. Big^{25} equals 1 if the observation belongs to one of the largest 25 banks in terms of asset size, and 0 otherwise in Equation (2a). Big^{3} (Big^{4-25}) equals 1 if the observation belongs to one of the largest three (four to 25) banks in terms of asset size, and 0 otherwise in Equation (2b).[15] We discuss the rationale for each of the variables in Appendix B.

Convexity serves as an exclusion restriction in our two-stage estimation procedure because, while both finance theory and empirical evidence suggest a direct link between tax convexity and derivative usage (Nance et al. 1993; Tufano 1996; Géczy et al. 1997; Graham and Rogers 2002), it is unlikely that tax convexity and bond spread are directly linked.[16]

Measurement of Bond Yield Spread

Following Elton et al. (2001), we use a tax-adjusted measure of bond yield spread that eliminates distortions stemming from mismatches of duration and convexity between bank bonds and benchmark Treasury bonds.[17] We present the details of this technique in Appendix C.

Measurement of Derivative Exposure

Ideally, a measure of derivative exposure disaggregates notional and fair values by long and short positions taken, class of instrument, time to maturity, degree of leverage, and level of complexity (Hull 1993; Wong 2000). However, neither SFAS 119 nor SFAS 133 requires disclosures detailed enough to obtain an ideal measure of derivative exposure. Wong (2000) suggests a comprehensive measure of derivative exposure within the constraints of the limited available disclosures. This measure includes the size of the portfolio exposed to market risk (measured as the notional value of long derivatives net of short derivatives) and the delta of the portfolio (i.e., the rate of change of the portfolio value with respect to a change in the underlying risk factor) because derivative exposure is driven by both these variables.

---

[14] Relevant studies from the finance literature are cited in Appendix B in the discussion of variables used in the multinomial logit model.

[15] Equation (2a) does not include separate, binary size variables for the largest three and largest four to 25 banks because none of the largest three banks is in the *Nonuser* or *Hedger* category.

[16] Imposing an exclusion restriction in the Heckman procedure is critical because the absence of such a restriction can lead to severe multicollinearity problems in the second-stage estimation (Francis et al. 2010).

[17] The tax and accounting literatures identify various settings where taxes are impounded in the prices of assets with tax-favored (or tax-disfavored) returns (Shackelford and Shevlin 2001). Holders of corporate bonds pay state and local taxes on interest payments, whereas holders of Treasury bonds do not. Therefore, the asymmetric tax treatment of corporate and Treasury bond interest payments likely results in corporate bonds offering higher pre-tax return to yield the same after-tax return as Treasury bonds.

Our measures of derivative exposure, which account for the complementary relation among notional amounts, fair values, and past changes in the underlying risk factors (e.g., interest rates), are based on Wong (2000). For hedging derivatives, we measure portfolio size as the net long notional amount (*IrHdg* and *FxHdg*) because banks typically disclose long and short positions in such derivatives when providing information about their hedging activities.[18] However, more than half the sample banks with trading derivatives do not disclose long and short positions in such derivatives separately. For these banks, we estimate the extent of the net exposure to market risk using the technique employed by the Office of the Comptroller of the Currency.[19] We compute the deltas of the portfolio for each group of derivatives (*DeltaIrHdg, DeltaFxHdg, DeltaIrTrd*, and *DeltaFxTrd*) as the change in fair value of the derivative portfolio divided by the change in the underlying market risk factor (interest rate or foreign exchange rate) during the fiscal year.[20,21]

Models for Testing Hypotheses 3 and 4

The bond-pricing models provide evidence on how bond investors perceive and price accounting measures of derivative exposures before and after SFAS 133, but do not directly test whether and how SFAS 133 altered the risk-offsetting relation between derivative and non-derivative exposures. Studies that examine how derivative exposures covary with other exposures typically examine the association between fair value gains/losses from derivatives and fair value gains/losses from non-derivative exposures (e.g., Venkatachalam 1996; Liu et al. 2004). Following Liu et al. (2004), we (1) regress the quarterly change in the net fair value of hedging derivatives ($\Delta DerHedging$) on the quarterly change in fair value of net non-derivative trading assets ($\Delta NonDerTrading$), and (2) regress the quarterly change in the net fair value of trading derivatives ($\Delta DerTrading$) on the quarterly change in fair value of net non-trading assets ($\Delta NonTrading$), respectively. A negative (positive) association of the dependent variable with $\Delta NonDerTrading$ and $\Delta NonTrading$ suggests that derivatives hedge (add to) the risks from non-derivative trading and non-trading exposures, respectively. We estimate the following models for the periods before and after SFAS 133:

$$\Delta DerHedging_{it} = \gamma_0 + \gamma_1 High_{it}^{Hdg} + \gamma_2 \Delta NonDerTrading_{it} + \gamma_3 High_{it}^{Hdg} * \Delta NonDerTrading_{it} + \gamma_4 \Delta NonTrading_{it} + \gamma_5 High_{it}^{Hdg} * \Delta NonTrading_{it} + e_{it}. \tag{3}$$

$$\Delta DerTrading_{it} = \lambda_0 + \lambda_1 High_{it}^{Trd} + \lambda_2 \Delta NonDerTrading_{it} + \lambda_3 High_{it}^{Trd} * \Delta NonDerTrading_{it} + \lambda_4 \Delta NonTrading_{it} + \lambda_5 High_{it}^{Trd} * \Delta NonTrading_{it} + e_{it}. \tag{4}$$

where $High^{Hdg}$ ($High^{Trd}$) equals 1 if the notional amount of derivatives classified as hedging (trading) scaled by beginning book value of equity is above the sample median for the year, and 0 otherwise. All other variables are scaled by beginning book value of equity. We interact $High^{Hdg}$ and $High^{Trd}$ with the fair value changes in non-derivative positions to control for potential differences in the covariation of derivative and non-derivative exposures across banks that hedge more

---

[18] For the few sample banks that hold basis swaps, we add the notional values of the basis swaps to the absolute difference between the notional values of long and short derivatives.

[19] In its derivative reports, the Office of the Comptroller of the Currency estimates the extent of non-offset derivative exposure to risk by dividing the absolute value of the net fair value of the derivative portfolio by the absolute value of the gross fair value of the derivative portfolio that is exposed to that specific risk.

[20] The terms of the interest rates used in the calculation of interest rate derivative portfolio deltas are based on the weighted average maturity of interest rate derivatives estimated using Call Report data. Then we compute the yield on a synthetic Treasury bond with the same maturity using the parametric yield curve specification suggested by Gurkaynak et al. (2006).

[21] The changes in foreign exchange rates are based on the Nominal Broad Dollar Index released by the Federal Reserve.

versus less and banks that trade more versus less, respectively, because banks with larger portfolios of derivatives are more likely to be impacted by SFAS 133. As predicted in H3, we expect the coefficient on Δ*NonDerTrading* in Equation (3) (i.e., γ_2 for banks with less and $\gamma_2 + \gamma_3$ for banks with more hedging derivatives) to be more negative after SFAS 133 because we expect hedging derivatives to comprise more effective hedges on average after SFAS 133. We also expect the coefficient on Δ*NonTrading* in Equation (4) (i.e., λ_4 for banks with less and $\lambda_4 + \lambda_5$ for banks with more trading derivatives) to be more negative (or less positive) as predicted in H4 because we expect SFAS 133 to increase the weight of derivative and non-derivative positions with offsetting exposures in the trading portfolio.

IV. EVIDENCE

Sample Selection

The primary criteria for inclusion in the sample are use of derivative instruments and availability of bond issue data. To ensure that financial data are available at the time of bond issuance, we match the Call Report (FR Y-9C) and 10-K data with the data for the first subordinated bond issue after the date of the 10-K filing. Consistent with the prior bond-pricing literature, our bond sample consists of subordinated bonds because subordinated bond spreads' greater sensitivity to credit and market risks facilitates more powerful tests of the relation between bank bond spread and accounting risk measures. We obtain bond issue data from the Mergent Fixed Income Securities Database (FISD).

We restrict our sample period to two years before and two years after SFAS 133 in order to mitigate the possibility that our results are driven by factors other than SFAS 133.[22] Using the Bank Compustat Database, we first identify 710 banks including both national commercial banks (SIC 6021) and state commercial banks (SIC 6022) that were publicly traded between 1999 and 2003. Using the Federal Reserve Bank of Chicago's Bank Regulatory Database, we identify 165 of these 710 banks as derivative users. We eliminate 24 of the 165 user banks due to lack of derivative disclosures in their 10-Ks, resulting in a final sample of 141 banks. We retain only one subordinated bond issue per bank in each year. These criteria yield a sample of 270 bond issues for the pre-SFAS 133 period and 265 bond issues for the post-SFAS 133 period.

Descriptive Statistics

We present descriptive statistics on the composition of sample bank derivative portfolios based on gross notional amounts in Panels A through D of Table 1. All 141 sample banks hold hedging derivatives, whereas only 52 (67) banks hold trading derivatives before (after) SFAS 133. As expected, banks carry mostly interest rate and foreign exchange derivatives. As presented in Panel A (Panel C), interest rate derivatives constitute the largest portion of derivatives, comprising 92.5 (67.4) percent and 88.3 (73.9) percent of hedging (trading) derivatives before and after SFAS 133, respectively. The weight of foreign exchange derivative contracts is higher in trading derivatives compared to hedging derivatives because the former include derivatives sold to customers that manage foreign exchange risk exposure. Swaps are the primary component of interest rate derivatives classified as hedging and as trading, whereas forwards are most prevalent among foreign exchange derivatives (Panels B and D, respectively).

Table 2 reports descriptive statistics for the dependent and independent variables employed in the bond-pricing tests. Tax-adjusted bond spread has a sample mean (median) of 70 (62) basis

[22] The tests for the pre-SFAS 133 (post-SFAS 133) period are based on financial statement data for fiscal years ending December 31, 1999 and 2000 (2002 and 2003). We exclude the year 2001, the first effective year of SFAS 133, because the annual change in fair values of hedging and trading derivatives from year 2000 to 2001 is distorted by reclassification of some derivatives into trading and hedging by SFAS 133.

TABLE 1

Composition of Derivative Portfolios Based on Gross Notional Amounts

Panel A: Composition of Hedging Derivatives Based on the Type of Underlying Risk Factor

	Pre-SFAS 133 (1999–2000)			Post-SFAS 133 (2002–2003)		
	No. of User Banks	Mean	Median	No. of User Banks	Mean	Median
Interest Rate Derivatives	139	0.925	1.000	139	0.883	1.000
Foreign Exchange Derivatives	26	0.057	0.000	22	0.065	0.000
Equity Derivatives	15	0.018	0.000	19	0.052	0.000
Commodity Derivatives	3	0.001	0.000	7	0.001	0.000

Panel B: Composition of Hedging Derivatives Based on the Type of Instrument

	Pre-SFAS 133 (1999–2000)			Post-SFAS 133 (2002–2003)		
	No. of User Banks	Mean	Median	No. of User Banks	Mean	Median
Interest Rate Derivatives						
Swaps	135	0.707	0.726	131	0.712	0.755
Options (Caps and Floors)	72	0.182	0.256	80	0.202	0.188
Forwards	18	0.036	0.000	11	0.013	0.000
Foreign Exchange Derivatives						
Swaps	5	0.001	0.000	2	0.000	0.000
Options	11	0.006	0.000	8	0.012	0.000
Forwards	19	0.050	0.000	16	0.009	0.000

Panel C: Composition of Trading Derivatives Based on the Type of Underlying Risk Factor

	Pre-SFAS 133 (1999–2000)			Post-SFAS 133 (2002–2003)		
	No. of User Banks	Mean	Median	No. of User Banks	Mean	Median
Interest Rate Derivatives	51	0.674	0.844	65	0.739	0.902
Foreign Exchange Derivatives	38	0.280	0.102	39	0.186	0.033

(continued on next page)

Panel C: Composition of Trading Derivatives Based on the Type of Underlying Risk Factor

	Pre-SFAS 133 (1999–2000)			Post-SFAS 133 (2002–2003)		
	No. of User Banks	Mean	Median	No. of User Banks	Mean	Median
Equity Derivatives	17	0.041	0.000	19	0.058	0.000
Commodity Derivatives	5	0.005	0.000	8	0.018	0.000

Panel D: Composition of Trading Derivatives Based on the Type of Instrument

	Pre-SFAS 133 (1999–2000)			Post-SFAS 133 (2002–2003)		
	No. of User Banks	Mean	Median	No. of User Banks	Mean	Median
Interest Rate Derivatives						
Swaps	50	0.476	0.367	54	0.494	0.408
Options (Caps and Floors)	30	0.152	0.098	35	0.178	0.102
Forwards	7	0.047	0.000	13	0.059	0.000
Foreign Exchange Derivatives						
Swaps	6	0.022	0.000	8	0.001	0.000
Options	19	0.048	0.000	26	0.031	0.000
Forwards	32	0.206	0.182	28	0.162	0.147

Statistics presented in Panels A and B are based on 141 banks with hedging derivatives.

Pre-SFAS 133 and post-SFAS 133 statistics presented in Panels C and D are based on 52 and 67 banks with trading derivatives, respectively.

Data presented in Panels A through D are as of December 31 of each year.

TABLE 2

Descriptive Statistics for Variables Employed in Bond-Pricing Tests

Variables	Mean	Median	Minimum	Maximum
Dependent Variable				
Tax-Adjusted Spread (%)	0.702	0.624	0.169	8.482
Derivative Exposure Variables Based on Portfolio Size				
IrHdg	0.085	0.079	0.005	0.603
FxHdg	0.001	0.000	0.000	0.068
IrTrd[a]	0.045	0.030	0.003	0.716
FxTrd[a]	0.012	0.002	0.000	0.452
Derivative Exposure Variables Based on Portfolio Delta				
DeltaIrHdg	0.004	0.004	0.000	0.025
DeltaFxHdg	0.001	0.000	0.000	0.004
DeltaIrTrd[a]	0.001	0.000	0.000	0.017
DeltaFxTrd[a]	0.001	0.000	0.000	0.010
Control Variables				
Maturity	11.785	10.000	1.000	30.000
Amount	17.443	17.076	6.451	24.635
Size	15.901	15.844	12.218	20.447
Insured	0.411	0.479	0.232	0.867
Estate	0.001	0.000	0.000	0.015
Npl	0.007	0.006	0.000	0.057
Gap	0.181	0.152	0.000	0.689
Roa	0.012	0.011	−0.061	0.077
Cap	0.076	0.078	0.033	0.237
NonDerTrd	0.042	0.019	0.000	0.266
CrRisk	0.000	0.000	0.000	0.001
LoanComm	0.064	0.056	0.000	0.465
Guarantee	0.007	0.000	0.000	0.153
CrDer	0.005	0.000	0.000	0.254
Strip	0.001	0.000	0.000	0.021

The sample represents 535 bank-year observations from 1999 to 2003 (excluding 2001) for 141 banks.

All control variables are as defined in Appendix A.

[a] Statistics presented are based only on banks with trading derivatives.

Variable Definitions:

IrHdg (*FxHdg*) = absolute value of net long/short notional value of interest rate (foreign exchange) hedging derivatives scaled by total assets;

IrTrd (*FxTrd*) = absolute value of net long/short notional value of interest rate (foreign exchange) trading derivatives scaled by total assets;

DeltaIrHdg (*DeltaFxHdg*) = portfolio delta of interest rate (foreign exchange) hedging derivatives; and

DeltaIrTrd (*DeltaFxTrd*) = portfolio delta of interest rate (foreign exchange) trading derivatives.

points. The mean ratio of *net* notional amount of interest rate hedging derivatives to total assets (*IrHdg*) is 8.5 percent, whereas the mean ratio of *gross* notional amount of interest rate hedging derivatives to total assets (unreported) is 12.8 percent. For banks with trading derivatives, the mean ratio of *net* notional amount of interest rate trading derivatives to total assets (*IrTrd*) is 4.5 percent, whereas the mean ratio of *gross* notional amount of interest rate trading derivatives to total assets (unreported) is 163.4 percent. These statistics indicate that, although the volume of trading derivatives is significantly greater than the volume of hedging derivatives, the size of net

exposure from hedging derivatives (*IrHdg*) is greater, on average, than the size of net exposure from trading derivatives (*IrTrd*), suggesting that banks take offsetting positions in trading derivatives to a great extent.

Unreported correlation analyses generally yield insignificant correlations between tax-adjusted spread and the variables of interest. One exception is *IrHdg*, which is negatively correlated with tax-adjusted spread both before ($r = -0.102$; $p = 0.094$) and after SFAS 133 ($r = -0.123$; $p = 0.046$). Also, the insignificant correlations (unreported) among the variables of interest suggest that multicollinearity between these variables of interest is unlikely to significantly affect our estimation results.

Results of the First-Stage Multinomial Logit Model

Table 3 contains results of the multinomial logit regressions for the *Nonuser* and the *Trader* categories. The results indicate that *Nonuser*, *Hedger*, and *Trader* banks systematically differ. Specifically, the results of the multinomial logit regression for *Nonuser* status indicate that banks less likely to use derivatives have a smaller asset portfolio (*Size*), lower probability of financial distress (*Lev*), lower interest rate risk exposure (*Gap*), lower credit risk (*Npl*), executive compensation plans with stronger incentives to take risk (*Incentive*), and less convex tax functions (*Convexity*), relative to *Hedger* banks. The results of the multinomial logit regression for *Trader* status indicate that banks more likely to use derivatives to a large extent and engage in derivatives trading have a larger asset portfolio (*Size*), lower probability of financial distress (*Lev*), lower net interest margin (*Nim*), and larger portfolio of commercial and industrial loans (*ComIndLoan*),

TABLE 3

Multinomial Logit Results

	Nonuser/Hedger			**Trader/Hedger**		
	Predicted Sign	**Coefficient**	**p-value**	**Predicted Sign**	**Coefficient**	**p-value**
Big^{25}	−	−0.099	0.133			
Big^{3}				+	0.211	0.008
Big^{4-25}				+	0.135	0.002
Size	−	−0.418	<0.001	+	0.365	0.004
Lev	−	−0.007	0.003	−	−0.005	0.025
Gap	−	−0.400	0.016	−/+	0.246	0.240
Nim	−/+	−3.079	0.007	−	−1.911	0.060
Npl	−	−1.284	0.002	−/+	0.921	0.084
Liquidity	−/+	0.609	0.247	+	0.035	0.162
Incentive	+	0.048	<0.001	−/+	0.033	0.181
Convexity	−	−0.005	0.053	−/+	0.002	0.014
ComIndLoan	−/+	0.152	0.042	+	0.542	0.072
Log Likelihood			−196.224			−368.357

The sample represents observations from 1999 to 2003 (excluding 2001).

The sample includes 710 publicly traded banks, 141 of which are derivative users. The tests are based on 10,920 bank-quarters: 8,720 observations belong to *Nonuser* banks, 1,268 observations belong to *Hedger* banks, and 932 observations belong to *Trader* banks.

All variables are as defined in Appendix B.

relative to *Hedger* banks. These findings are in line with risk management theory and the findings of prior empirical studies.[23]

Results of Tests of Hypotheses 1 and 2

The data set used in this study contains observations on multiple banks across multiple years. To estimate standard errors that are robust to both time-series and cross-sectional correlation, all results are based on standard errors clustered simultaneously by bank and by year.

Table 4, Panel A reports estimation results for the model presented in Equation (1). The coefficient signs on the control variables are generally consistent with those reported in prior studies. *Iml*, the Inverse Mills ratio obtained from the multinomial logit model, is statistically different from zero both before and after SFAS 133 ($p < 0.010$ and $p < 0.002$, respectively), underscoring the necessity to control for potential self-selection bias.

We find negative and significant associations between tax-adjusted spread and both accounting measures of exposure from interest rate hedging derivatives, *IrHdg* and *DeltaIrHdg*, before and after SFAS 133.[24] These findings are consistent with such derivatives being used for hedging purposes and the complementary relation between derivative notional amounts and fair values in capturing derivative exposures as pointed out by Wong (2000). To test H1 and H2, we test the equality of coefficients on derivative exposure measures across pre- and post-SFAS 133 periods for the largest 25 and the remaining banks. The results of these tests are presented in Table 4, Panel B. For both subsamples, as hypothesized in H1, the associations of tax-adjusted bond spread with *IrHdg* and *DeltaIrHdg* are more negative after SFAS 133 ($p = 0.018$ and $p = 0.015$ for the largest 25 banks; $p = 0.001$ and $p = 0.032$ for the remaining banks, respectively). These findings are consistent with derivative hedges that are more effective, on average, being classified as hedging after SFAS 133.

The insignificant coefficients on the interactions of *IrHdg*, *DeltaIrHdg*, *FxHdg*, and *DeltaFxHdg* with *Big*[25] before and after SFAS 133 (presented in Table 4, Panel A) suggest that the pricing of hedging derivatives in bond markets does not differ between the largest 25 banks and the remaining banks.[25]

The results in Table 4, Panel B also show that the associations between tax-adjusted spread and both *IrTrd* and *DeltaIrTrd* are more negative after SFAS 133 for the largest 25 banks ($p = 0.018$ and $p = 0.019$, respectively). Regarding the two foreign exchange derivative exposure measures (*FxTrd* and *DeltaFxTrd*), we observe a similar pattern only for *FxTrd* ($p = 0.089$). The negative coefficients on *IrTrd, DeltaIrTrd*, and *FxTrd* are consistent with bond markets, on average, viewing trading derivatives as being used to cross-hedge other risk exposures. The findings of greater negative coefficients on *IrTrd, DeltaIrTrd*, and *FxTrd* following SFAS 133 for the large banks are consistent with more economic hedges being classified as trading after SFAS 133 by these banks. These findings generally support H2. However, for the remaining banks, we find no support for H2, perhaps because these relatively smaller banks do not have much involvement in or exposures from trading derivatives. Also, except for supporting H2 for the largest 25 banks, H1 and H2 are not supported for foreign exchange rate derivatives, perhaps because banks' exposures

---

[23] The multinomial logit regression for *Nonuser* (*Trader*) status correctly classifies 89 percent (77 percent) of banks.

[24] When tax-adjusted bond spreads are decomposed into systematic risk and default risk premium components as suggested by Elton et al. (2001), we find that the negative association with spread is most pronounced for the systematic risk premium component, whereas the magnitude of the negative association with default risk premium is modest. The results of these tests are excluded for brevity.

[25] The negative and positive associations of tax-adjusted spread with accounting measures of derivative exposures and *Gap*, respectively, are in line with the findings of Schrand (1997) who documents that maturity gap (non-trading interest rate risk) is positively associated with higher perceived risk, and hedging derivatives decrease the maturity gap, thus reducing perceived risk.

TABLE 4

Relation between Bank Bond Spread and Derivatives before and after SFAS 133

Panel A: Association of Spread with Derivatives

	Predicted Sign	Pre-SFAS 133 Period		Post-SFAS 133 Period	
		Coefficient	p-value	Coefficient	p-value
Big^{25}	–	−0.231	0.050	−0.163	0.030
IrHdg	–	−1.827	0.001	−2.862	0.001
$Big^{25} * IrHdg$	−/+	−0.172	0.139	−0.310	0.208
DeltaIrHdg	–	−10.008	0.077	−17.656	0.048
$Big^{25} * DeltaIrHdg$	−/+	−1.901	0.124	−1.366	0.242
FxHdg	–	−0.660	0.383	0.726	0.206
$Big^{25} * FxHdg$	−/+	0.712	0.289	0.271	0.134
DeltaFxHdg	–	−0.981	0.447	1.325	0.110
$Big^{25} * DeltaFxHdg$	−/+	0.235	0.268	−0.451	0.327
IrTrd	−/+	0.469	0.177	0.119	0.165
$Big^{25} * IrTrd$	−/+	−1.900	0.002	−2.586	0.009
DeltaIrTrd	−/+	−0.161	0.239	0.766	0.342
$Big^{25} * DeltaIrTrd$	−/+	−5.989	0.060	−9.354	0.048
FxTrd	−/+	0.409	0.207	−0.090	0.192
$Big^{25} * FxTrd$	−/+	−1.441	0.001	−1.222	0.087
DeltaFxTrd	−/+	−0.312	0.317	0.795	0.280
$Big^{25} * DeltaFxTrd$	−/+	−1.566	0.233	−1.169	0.392
Maturity	+	0.088	0.032	0.075	0.001
Amount	–	−0.078	0.059	−0.060	0.075
Size	–	−0.054	0.009	−0.078	0.001
Insured	+	2.885	0.154	4.061	0.143
Estate	+	83.889	0.288	60.009	0.204
Npl	+	36.419	0.005	31.593	0.001
Gap	+	1.998	0.009	1.834	0.001
Roa	–	−92.278	0.001	−71.080	0.001
Cap	–	−29.368	0.001	−32.721	0.006

(continued on next page)

Panel A: Association of Spread with Derivatives

	Predicted Sign	Pre-SFAS 133 Period Coefficient	Pre-SFAS 133 Period p-value	Post-SFAS 133 Period Coefficient	Post-SFAS 133 Period p-value
NonDerTrd	+	2.800	0.007	2.599	0.013
CrRisk	+	10.385	0.244	3.501	0.277
LoanComm	−	−0.907	0.078	−1.282	0.080
Guarantee	−	0.091	0.260	−0.482	0.164
CrDer	−	−1.020	0.075	−0.866	0.161
Strip	−	−8.187	0.054	−5.588	0.090
Iml	−/+	−0.593	0.010	−0.736	0.002
Adjusted R^2			0.737		0.779
No. of Observations			270		265

Panel B: Tests of Hypotheses 1 and 2 Based on the Results Presented in Panel A

$$r_{it} = \beta_0 + \beta_1 Big_{it}^{25} + \beta_2 IrHdg_{it} + \beta_3 Big_{it}^{25} * IrHdg_{it} + \beta_4 DeltaIrHdg_{it} + \beta_5 Big_{it}^{25} * DeltaIrHdg_{it} + \beta_6 FxHdg_{it} + \beta_7 Big_{it}^{25} * FxHdg_{it} + \beta_8 DeltaFxHdg_{it} + \beta_9 Big_{it}^{25} * DeltaFxHdg_{it} + \beta_{10} IrTrd_{it} + \beta_{11} Big_{it}^{25} * IrTrd_{it} + \beta_{12} DeltaIrTrd_{it} + \beta_{13} Big_{it}^{25} * DeltaIrTrd_{it} + \beta_{14} FxTrd_{it} + \beta_{15} Big_{it}^{25} * FxTrd_{it} + \beta_{16} DeltaFxTrd_{it} + \beta_{17} Big_{it}^{25} * DeltaFxTrd_{it} + \beta_{18} Maturity_{it} + \beta_{19} Amount_{it} + \beta_{20} Size_{it} + \beta_{21} Insured_{it} + \beta_{22} Estate_{it} + \beta_{23} Npl_{it} + \beta_{24} Gap_{it} + \beta_{25} NonDerTrd_{it} + \beta_{26} Cap_{it} + \beta_{27} Roa_{it} + \beta_{28} CrRisk_{it} + \beta_{29} LoanComm_{it} + \beta_{30} Guarantee_{it} + \beta_{31} CrDer_{it} + \beta_{32} Strip_{it} + \beta_{33} Iml_{it} + \varepsilon_{it}$$

	Largest 25 Banks	p-value	Remaining Banks	p-value
Hypothesis 1	$\beta_2^{PRE} + \beta_3^{PRE} > \beta_2^{POST} + \beta_3^{POST}$	0.018	$\beta_2^{PRE} > \beta_2^{POST}$	0.001
	$\beta_4^{PRE} + \beta_5^{PRE} > \beta_4^{POST} + \beta_5^{POST}$	0.015	$\beta_4^{PRE} > \beta_4^{POST}$	0.032
	$\beta_6^{PRE} + \beta_7^{PRE} > \beta_6^{POST} + \beta_7^{POST}$	0.341	$\beta_6^{PRE} > \beta_6^{POST}$	0.297
	$\beta_8^{PRE} + \beta_9^{PRE} > \beta_8^{POST} + \beta_9^{POST}$	0.205	$\beta_8^{PRE} > \beta_8^{POST}$	0.227
Hypothesis 2	$\beta_{10}^{PRE} + \beta_{11}^{PRE} > \beta_{10}^{POST} + \beta_{11}^{POST}$	0.018	$\beta_{10}^{PRE} > \beta_{10}^{POST}$	0.238
	$\beta_{12}^{PRE} + \beta_{13}^{PRE} > \beta_{12}^{POST} + \beta_{13}^{POST}$	0.019	$\beta_{12}^{PRE} > \beta_{12}^{POST}$	0.307
	$\beta_{14}^{PRE} + \beta_{15}^{PRE} > \beta_{14}^{POST} + \beta_{15}^{POST}$	0.089	$\beta_{14}^{PRE} > \beta_{14}^{POST}$	0.372
	$\beta_{16}^{PRE} + \beta_{17}^{PRE} > \beta_{16}^{POST} + \beta_{17}^{POST}$	0.392	$\beta_{16}^{PRE} > \beta_{16}^{POST}$	0.274

(continued on next page)

The sample represents observations from 1999 to 2003 (excluding 2001) for 141 banks.
All p-values are based on one-tailed t-tests when coefficient signs are predicted, and based on two-tailed tests otherwise.
All variables are as defined in Table 2 and Appendix A.

from foreign exchange rate derivatives are not material enough to affect their bond spreads.

Among the variables that proxy for exposures from off-balance sheet instruments (other than derivatives) that are potentially used for hedging, *LoanComm* and *Strip* are negatively and significantly related to tax-adjusted bond spread before and after SFAS 133 ($p = 0.078$ and $p = 0.080$ for *LoanComm*; $p = 0.054$ and $p = 0.090$ for *Strip*, respectively). *CrDer* is also negatively related to tax-adjusted spread, but only prior to SFAS 133 ($p = 0.075$).

The results reported Table 4, Panel A indicate that the association between bond spread and accounting measures of derivative exposures is driven mainly by interest rate derivatives. In unreported tests, we find support for both H1 and H2 based on different types of interest rate derivative instruments (i.e., interest rate swaps and interest rate options).

An examination of sample bank 10-Ks and Call Reports reveal that 97 of the 141 sample banks had to transfer derivatives previously classified as hedging to trading because of the various provisions of SFAS 133, indicating that SFAS 133 had direct effects on the classification of derivatives for a majority of the sample banks. Our hypotheses still hold for the banks that did not have to reclassify any derivatives. The more stringent criteria imposed by SFAS 133 (rather than the relatively lax criteria before SFAS 133) assure that derivatives classified as hedging in the post-SFAS 133 period are, in fact, highly effective hedging derivatives. In addition, as argued by the FASB (SFAS 133, ¶ 238), SFAS 133's detailed documentation requirements likely improve the monitoring of hedging activities and result in less uncertainty in the financial statements, leading to lower bank bond spreads for all derivative-user banks. Overall, these results support the view that banks benefit from the more credible derivative-related information provided under SFAS 133.

Economic Effects of SFAS 133 on Bond Spreads

The findings from the bond-pricing tests indicate that the association between the accounting measures of derivative exposures and bond spread is driven mainly by interest rate derivatives. To provide some indication of the economic effects of interest rate derivatives before and after SFAS 133, we estimate their effects on bond spread using the estimated coefficients reported in Panel A of Table 4 and descriptive statistics reported in Table 2. All estimates for the periods before and after SFAS 133 are based on mean observations for the entire sample period.

Our results indicate that banks with average values of *IrHdg* and *DeltaIrHdg* have bond spreads that are 19.5 basis points lower in the pre-SFAS 133 period and 30.8 basis points lower in the post-SFAS 133 period than banks with no exposure from interest rate hedging derivatives.[26] Similarly, banks with average values of *IrTrd* and *DeltaIrTrd* have bond spreads that are 5.8 basis points lower in the pre- and 9.9 basis points lower in the post-SFAS 133 period than banks with no exposure from interest rate trading derivatives.[27]

Results of Tests of Hypotheses 3 and 4

Panels A through D of Table 5 present post-SFAS 133 and pre-SFAS 133 Spearman correlations (above and below the diagonal, respectively) for the variables used in the exposure co-variation tests. First, we partition the entire sample into banks with large and small portfolios of hedging derivatives.[28] Panels A and B display correlations among the variables used to estimate

---

26 We use the weighted average of the coefficients for the largest 25 banks and other banks when estimating the effects of hedging derivatives because the effect of hedging derivatives on bond spread is significantly negative for both groups of banks.

27 We estimate the effects of trading derivatives only for the largest 25 banks because our results indicate that the effect of trading derivatives on bond spread is significantly negative for only these banks.

28 We consider a bank to have a large (small) portfolio of hedging derivatives if the notional amount of hedging derivatives scaled by beginning book value of equity is above (below) the sample median in the pre-SFAS 133 period.

TABLE 5

Association of Derivative and Non-Derivative Exposures before and after SFAS 133
Correlation Analyses

Panel A: Correlations for High Users of Hedging Derivatives

	ΔDerTrading	*ΔDerHedging*	*ΔNonDerTrading*	*ΔNonTrading*
ΔDerTrading		−0.049 (0.245)	−0.103 (0.015)	0.044 (0.303)
ΔDerHedging	−0.019 (0.655)		−0.051 (0.229)	−0.284 (0.001)
ΔNonDerTrading	−0.062 (0.140)	0.033 (0.431)		−0.028 (0.503)
ΔNonTrading	−0.057 (0.177)	−0.227 (0.001)	0.046 (0.272)	

Panel B: Correlations for Low Users of Hedging Derivatives

	ΔDerTrading	*ΔDerHedging*	*ΔNonDerTrading*	*ΔNonTrading*
ΔDerTrading		0.069 (0.102)	−0.068 (0.106)	−0.038 (0.367)
ΔDerHedging	0.005 (0.915)		0.033 (0.441)	−0.175 (0.001)
ΔNonDerTrading	−0.044 (0.297)	−0.009 (0.830)		−0.002 (0.957)
ΔNonTrading	0.027 (0.528)	−0.155 (0.001)	0.030 (0.474)	

Panel C: Correlations for High Users of Trading Derivatives

	ΔDerTrading	*ΔDerHedging*	*ΔNonDerTrading*	*ΔNonTrading*
ΔDerTrading		−0.019 (0.758)	−0.128 (0.039)	0.044 (0.479)
ΔDerHedging	−0.031 (0.657)		−0.034 (0.587)	−0.251 (0.001)

(continued on next page)

Panel C: Correlations for High Users of Trading Derivatives

	ΔDerTrading	*ΔDerHedging*	*ΔNonDerTrading*	*ΔNonTrading*
ΔNonDerTrading	−0.091 (0.196)	0.089 (0.207)		0.050 (0.418)
ΔNonTrading	0.021 (0.772)	−0.204 (0.004)	−0.007 (0.925)	

Panel D: Correlations for Low Users of Trading Derivatives

	ΔDerTrading	*ΔDerHedging*	*ΔNonDerTrading*	*ΔNonTrading*
ΔDerTrading		0.090 (0.147)	−0.073 (0.238)	−0.022 (0.718)
ΔDerHedging	0.045 (0.526)		0.035 (0.570)	−0.190 (0.002)
ΔNonDerTrading	−0.056 (0.426)	−0.061 (0.389)		0.040 (0.522)
ΔNonTrading	0.078 (0.273)	−0.148 (0.035)	−0.005 (0.940)	

Correlation coefficients in Panels A and B are based on financial data from 1999 to 2003 (excluding 2001) for 141 banks.

Correlation coefficients in Panels C and D are based on financial data for 52 banks in the pre-SFAS 133 and 67 banks in the post-SFAS 133 period.

Post-SFAS 133 and pre-SFAS 133 Spearman correlations are presented above and below the diagonal, respectively.

Significance levels in parentheses are based on two-tailed tests.

Variable Definitions:

ΔDerTrading = quarterly change in the fair value of net trading derivative assets scaled by beginning book value of equity;
ΔDerHedging = quarterly change in the fair value of net hedging derivative assets scaled by beginning book value of equity;
ΔNonDerTrading = quarterly change in the fair value of net non-derivative trading assets scaled by beginning book value of equity; and
ΔNonTrading = quarterly change in the fair value of net non-trading assets scaled by beginning book value of equity.

Equation (3). As presented in Panel A, for high-hedger banks, the correlation between *ΔDerHedging* and *ΔNonTrading* is more negative after SFAS 133 (i.e., it changes from −0.227 to −0.284), providing preliminary evidence in support of H3. Although not as sizable, the change in the correlation coefficient (from −0.155 to −0.175) is in the same direction for banks with small portfolios of hedging derivatives as presented in Panel B.

Panels C and D of Table 5 display correlations among the variables used in estimating Equation (4) for banks with large and small portfolios of trading derivatives, respectively. The negative correlations between *ΔDerTrading* and *ΔNonDerTrading* presented in Panels C and D provide evidence in support of trading derivatives offsetting exposures from non-derivative trading positions. For high-trader banks, which primarily consist of large banks, the correlation between *ΔDerTrading* and *ΔNonDerTrading* is more negative after the adoption of SFAS 133 ($r = -0.128$; $p = 0.039$) compared to the pre-SFAS 133 period ($r = -0.091$;$p = 0.196$), consistent with more economic hedges being classified as trading after SFAS 133. A similar trend also exists for low-trader banks. However, the correlation coefficients are smaller and insignificant ($r = -0.056$; $p = 0.426$ and $r = -0.073$; $p = 0.238$ before and after SFAS 133, respectively).

The insignificant correlation coefficients between *ΔDerHedging* and *ΔNonDerTrading* (*ΔDerTrading* and *ΔNonTrading*) observed in Panels A through D indicate that banks do not cross-hedge trading (non-trading) positions with hedging (trading) derivatives. These findings are consistent with Liu et al. (2004), who find no significant cross-hedging between trading exposures and non-trading exposures.

The results from estimating Equation (3) that are reported in Panel A of Table 6 are mostly consistent with the correlation analyses. Panel C presents the results of the tests of H3 for banks with large and small portfolios of hedging derivatives. The coefficient on *ΔNonTrading* is more negative after SFAS 133 for both groups of banks ($p = 0.030$ and $p = 0.078$, respectively), providing further evidence that derivatives classified as hedging consist of hedges that meet stricter hedge effectiveness criteria after SFAS 133.

Results from estimating Equation (4) are reported in Panel B of Table 6. Panel C presents the results of the tests of H4 for banks with large and small portfolios of trading derivatives. The negative and significant coefficient on *ΔNonDerTrading* for both groups of banks before and after SFAS 133 suggests that trading derivatives cross-hedge non-derivative trading exposures. Also, the coefficient on *ΔNonDerTrading* is more negative subsequent to SFAS 133 for banks with large portfolios of trading derivatives ($p = 0.054$), providing support for H4. More specifically, this finding suggests that the extent of cross-hedging between trading derivatives and non-derivative trading positions is greater after the adoption of SFAS 133, consistent with more economic hedges that do not qualify as accounting hedges being classified as trading subsequent to SFAS 133. However, for banks with small portfolios of trading derivatives, the coefficients on *ΔNonDerTrading* are not reliably different before and after SFAS 133 ($p = 0.151$).

Sensitivity Analyses and Additional Considerations

We conduct several sensitivity analyses to assess the robustness of our findings. First, we re-run our tests after deleting the three largest banks (J.P. Morgan Chase, Bank of America, and Citibank) that dominate the derivatives market during our sample period. We continue to find support for all hypotheses. Second, a potential explanation for our findings is that the change in the yield curve coincides with the adoption of SFAS 133. The flat yield curve during the pre-SFAS 133 period (1999–2000) became steeply upward sloped in 2001. It remained so through the end of 2004 and flattened in the 2005–2006 period. In order to address the possibility that our results are driven by the different-shaped yield curves in the pre- and post-SFAS 133 periods rather than by SFAS 133, we re-test our hypotheses using the years 2005–2006 as the post-SFAS 133 period. This specification ensures that both pre- and post-SFAS 133 period yield curves are flat. We

TABLE 6

Association of Derivative and Non-Derivative Exposures before and after SFAS 133 Regression Analyses

Panel A: Hedging Derivatives and Non-Derivative Exposures

	Pre-SFAS 133 Period		Post-SFAS 133 Period	
	Coefficient	p-value	Coefficient	p-value
$High^{Hdg}$	0.010	0.183	0.009	0.253
$\Delta NonDerTrading$	0.010	0.462	−0.007	0.358
$High^{Hdg} * \Delta NonDerTrading$	0.021	0.657	0.003	0.405
$\Delta NonTrading$	−0.108	0.022	−0.144	0.006
$High^{Hdg} * \Delta NonTrading$	−0.051	0.005	−0.081	0.041
Adjusted R^2		0.182		0.218
No. of Observations		1,048		1,014

Panel B: Trading Derivatives and Non-Derivative Exposures

	Pre-SFAS 133 Period		Post-SFAS 133 Period	
	Coefficient	p-value	Coefficient	p-value
$High^{Trd}$	−0.003	0.284	0.012	0.176
$\Delta NonDerTrading$	−0.056	0.038	−0.070	0.008
$High^{Trd} * \Delta NonDerTrading$	−0.030	0.021	−0.071	0.064
$\Delta NonTrading$	−0.034	0.412	0.093	0.222
$High^{Trd} * \Delta NonTrading$	0.041	0.255	0.002	0.334
Adjusted R^2		0.101		0.134
No. of Observations		404		510

(continued on next page)

Panel C: Tests of Hypotheses 3 and 4 Based on the Results Presented in Panels A and B

	High Users of Hedging Derivatives	p-value	Low Users of Hedging Derivatives	p-value
Hypothesis 3	$\gamma_4^{PRE} + \gamma_5^{PRE} > \gamma_4^{POST} + \gamma_5^{POST}$	0.030	$\gamma_4^{PRE} > \gamma_4^{POST}$	0.078

	High Users of Trading Derivatives	p-value	Low Users of Trading Derivatives	p-value
Hypothesis 4	$\lambda_2^{PRE} + \lambda_3^{PRE} > \lambda_2^{POST} + \lambda_3^{POST}$	0.054	$\lambda_2^{PRE} > \lambda_2^{POST}$	0.151

All results are based on standard errors clustered simultaneously by bank and by year.
All p-values are based on two-tailed t-tests.
All other variables are as defined in Table 5.

Variable Definition:
$High^{Hdg}$
($High^{Trd}$) = 1 if the notional amount of hedging (trading) derivatives scaled by beginning book value of equity is above sample median for the year, and 0 otherwise.

continue to find strong support for our hypotheses, suggesting that the change in the yield curve in 2001 is not driving our results. Third, as mentioned earlier, we estimate the size of the net long/short positions in trading derivatives for some banks. These estimates are likely to contain error. Therefore, we re-run our tests using gross (long plus short) notional values for trading derivatives. Our results regarding the effect of SFAS 133 still hold. Fourth, we repeat our tests after excluding the top and bottom 1 percent of the distribution of each variable to assess the sensitivity of our findings to extreme observations. We continue to observe results consistent with our primary findings. Last, we find similar results for the exposure covariation tests of H3 and H4 when we scale all variables by beginning total assets instead of beginning book value of equity as we do in our primary tests.

Another potential alternative explanation for the difference between the risk relevance of derivatives before and after SFAS 133 is the possibility of a change in banks' derivative strategies. However, from a cost-benefit standpoint, we believe that there are a number of reasons for banks not to change their derivative strategies following SFAS 133. First, derivatives are efficient and low-cost risk management tools for banks. In the absence of derivatives, managing interest rate risk would require significant amounts of capital investment. In addition, derivative contracts can be customized, depending on the risk being hedged. Thus, the benefits of managing interest rate risk via efficient and low-cost risk management tools likely outweigh the costs of using alternative hedging strategies. Second, a major concern about SFAS 133 expressed by banks is the increased volatility likely to be introduced by SFAS 133 to banks' earnings.[29] However, banks can potentially offset this volatility in a cost-effective way by using other tools for smoothing income (such as loan loss provisions) without changing their derivative strategies (Wahlen 1994; Liu and Ryan 2006; Kanagaretnam et al. 2010). In fact, Kilic et al. (2010) document that banks that are likely to be most impacted by SFAS 133 have increased their reliance on loan loss provisions for income smoothing in the post-SFAS 133 period.

Finally, we note that, although the volume of trading derivatives grew rapidly during our sample period,[30] this rapid growth is unlikely to drive our results regarding the effects of SFAS 133 for a number of reasons. First, our hypotheses are stated and models designed to test whether the pricing of a unit of derivative exposure changed with SFAS 133, regardless of the amount of total exposure. Second, the growth trend cannot be generalized to all banks; it is driven mostly by the largest seven derivative-dealer banks (Office of the Comptroller of the Currency Report on Derivatives Activities 1995–2006). We continue to find support for all our hypotheses when we re-run our tests after deleting these seven banks from the sample. Third, the growth trend started in the early 1990s and has continued since then. Thus, there is no evidence of a shift in the trend that coincides with the adoption of SFAS 133.

V. SUMMARY AND CONCLUSION

SFAS 133 is a widely debated and controversial accounting standard that replaced the incomplete and inconsistent accounting guidance for derivatives that existed prior to its adoption. We examine how SFAS 133 affected the risk relevance of accounting measures of bank derivative exposures to bond investors. We hypothesize and find that interest rate hedging derivatives are more negatively associated with bond spreads after SFAS 133. These findings are consistent with

[29] SFAS 133 was expected to introduce more volatility to banks' earnings because it requires more economic hedges to be classified as trading with gains/losses recognized in income immediately and any hedge ineffectiveness to be recognized in income immediately

[30] The annual growth rates of derivatives classified as trading are 35.71 percent in 1996, 29.46 percent in 1997, 53.65 percent in 1998, 3.25 percent in 1999, 17.88 percent in 2000, 11.92 percent in 2001 (the year of adoption of SFAS 133), 21.32 percent in 2002, 28.76 percent in 2003, 28.61 percent in 2004, 5.38 percent in 2005, and 11.52 percent in 2006.

SFAS 133 restricting derivatives classified as hedging to only those instruments that serve as highly effective and documented derivative hedges. For the largest 25 banks, which are extensively involved in derivatives trading, we find that interest and foreign exchange rate trading derivatives are more *negatively* associated with bond spreads after SFAS 133, consistent with more economic hedges being classified as trading after SFAS 133. To provide further insights, we also examine how SFAS 133 altered the association of exposures from derivatives with non-derivative exposures. Consistent with our results from the risk-relevance tests, the findings from the exposure covariation tests indicate that the offsetting effects between hedging derivatives and non-trading exposures, and between trading derivatives and non-derivative trading exposures, are greater after SFAS 133. Taken together, these findings indicate that, despite its limitations, SFAS 133 has improved the transparency and the monitoring of the risk implications of derivatives.

APPENDIX A
CONTROL VARIABLES USED IN THE BOND-PRICING MODELS

Variable Name	Data Source	Definition and Explanation
Bond Characteristics		
Maturity	Mergent FISD	Natural logarithm of the bond's years to maturity. Bonds with longer maturities (*MATURITY*) are expected to have higher spreads due to lower liquidity (Hancock and Kwast 2001; Covitz et al. 2000).
Amount	Mergent FISD	Natural logarithm of the size of the bond issue in millions of dollars. Bonds issued in larger amounts are expected to have lower spreads due to higher liquidity benefits (Hancock and Kwast 2001; Morgan and Stiroh 2001) and due to economies of scale in underwriting (Sengupta 1998).
Bank Characteristics		
Size	FR 9-YC	Natural logarithm of the bank's total assets. Larger banks are less likely to default because they are better diversified and are more likely to benefit from the "too-big-to-fail" policy formalized by the Comptroller of the Currency (Flannery and Sorescu 1996; Morgan and Stiroh 2001). Also, their bonds are more liquid and, therefore, are traded at lower spreads over Treasury bonds (Flannery and Sorescu 1996).
Insured	Bank Compustat	Ratio of total insured deposits to total deposits. Banks with larger insured deposits are subject to less market monitoring and discipline (Billett et al. 1998) and, therefore, shift to greater use of insured deposits as their financial condition deteriorates (Billett et al. 1998; Jordan 2000). As a result, banks with higher insured deposits are perceived to have greater potential for moral hazard and, therefore, higher bond spreads (Jagtiani et al. 2002).
Estate	FR 9-YC	Real estate acquired in satisfaction of debt previously contracted scaled by total assets. Creditors may view higher property collected on foreclosed real estate loans as a signal of lower loan quality and higher credit risk, and may demand higher bond spread (Morgan and Stiroh 2001).

(continued on next page)

Variable Name	Data Source	Definition and Explanation
Npl	FR 9-YC	Sum of nonaccrual loans and accruing loans past due 90 days or more scaled by total assets. Higher nonaccrual and past due loans signal lower loan quality and higher loan losses. Higher loan losses reduce the capital base more and require greater charges against income through higher provisions for loan losses, thereby increasing bank default risk (Flannery and Sorescu 1996; Morgan and Stiroh 2001).
Gap	10-K	Absolute value of the difference between bank non-trading assets and non-trading liabilities maturing or repricing within one year scaled by total assets. Higher maturity mismatch between non-trading assets and liabilities increases a bank's non-trading interest rate risk exposure and overall risk (Flannery and James 1984; Flannery and Sorescu 1996; Morgan and Stiroh 2001).
NonDerTrd	10-K	Fair value of non-derivative trading assets scaled by total assets. Trading activities are associated with higher bond spreads due to potential agency risks associated with such activities. That is, banks may more easily assume trading risks at levels that are suboptimal for bondholders, since trading positions can be easily changed or concealed with leverage. Such uncertainty about trading activities gives rise to agency risks (Morgan and Stiroh 2001).
Cap	FR Y9-C	Ratio of book value of equity to risk-weighted assets. A bank's assets differ in terms of riskiness. Some assets (e.g., commercial loans) are considered riskier than others (e.g., real estate loans), and are given higher risk weights in the calculation of regulatory capital. Bank capital is a cushion for potential losses. Thus, *Cap* will be lower for banks with riskier assets. *Cap* jointly captures the asset risk composition and capital adequacy. Higher capital is likely to result in creditors demanding lower spread.
Roa	FR Y9-C	Annual net income scaled by total assets. Higher profitability is associated with lower likelihood of default and bond spread (Flannery and Sorescu 1996; Morgan and Stiroh 2001).
CrRisk	FR Y9-C	Charge-offs incurred on derivative contracts during the year (i.e., losses incurred when the counterparty fails to meet the settlement obligations of the derivative contract) scaled by total assets. Our derivative exposure measures proxy market risk exposure. However, derivatives also expose the bank to credit risk. The charge-offs incurred on derivatives are expected to increase bond spread.
Off-Balance Sheet Hedges other than Market Risk Derivatives		
LoanComm	10-K	Amount of loans for which official promises to lend are conveyed scaled by total assets. Banks sometimes use forward loan sales commitments as economic hedges of commitments to originate loans (Office of the Comptroller of the Currency 2005) because fair values of these commitments are offsetting.

(continued on next page)

Variable Name	Data Source	Definition and Explanation
Guarantee	10-K	On- and off-balance sheet guarantees and insurance contracts that back financial obligations scaled by total assets. Banks enter into financial guarantee and insurance contracts that back off-balance sheet instruments or financial obligations such as outstanding securities and loans.
CrDer	10-K	Notional amount of credit derivatives on which the bank is the beneficiary scaled by total assets. In order to hedge their credit risks (losses) from the loans they originate, banks sometimes enter into credit derivative contracts with the bank as the beneficiary. Such contracts constitute guarantees against a bank's credit losses from other parties and are expected to reduce bank credit risk and thus bond spread.
Strip	10-K	Fair value of principal-only strips scaled by total assets. Principal-only strips do not qualify as derivative instruments under the requirements of SFAS 133 (paragraph 14), but these instruments are used to hedge prepayment risk especially from mortgage servicing rights, which is driven by decreasing interest rates.

APPENDIX B
VARIABLES USED IN THE FIRST-STAGE MULTINOMIAL LOGIT MODEL

Variable Name	Data Source	Definition and Explanation
Big_{it}^{25}	FR Y-9C	Dummy variable that equals 1 if the bank is one of largest 25 banks in the United States, and 0 otherwise. Holdings of derivatives in the United States are concentrated in the largest 25 banks, whereas the remaining holdings (particularly for hedging purposes) are still economically significant for the remaining user banks because their asset portfolios are remarkably smaller than those of the largest 25 banks. In order to better capture the effect of large banks, we include discrete size variables.
$Big_{it}^{3}(Big_{it}^{4-25})$	FR Y-9C	Dummy variable that equals 1 if the bank is one of largest three (4 to 25) banks in the United States, and 0 otherwise. Among the largest 25 banks, the largest three banks (J.P. Morgan Chase, Bank of America, and Citigroup) dominate derivatives activity (Office of the Comptroller of the Currency Derivatives Report, 1998−2003).
Size	FR Y-9C	Natural logarithm of total assets. A number of studies (e.g., Booth et al. 1984; Koppenhaver 1990; Kim and Koppenhaver 1992; Gunther and Siems 2002) document that bank size is positively related to derivative use because large banks can more easily bear costs associated with implementing derivative trading and hedging programs such as the costs of personnel training and internal control systems.

(continued on next page)

Variable Name	Data Source	Definition and Explanation
Lev	FR Y-9C	Ratio of risk-weighted assets to book value of equity. *Lev* jointly captures the probability of financial distress (Nance et al. 1993; Géczy et al. 1997; Sinkey and Carter 2000) and balance sheet composition in terms of asset riskiness. Banks with higher leverage have higher probability of financial distress, and are more likely to use derivatives to reduce the variance of bank value. Also, banks holding assets with higher credit risks (i.e., higher risk-weighted assets and therefore higher *Lev*) are more likely to use derivatives to hedge other exposures (Schrand and Unal 1998). On the other hand, a bank with low capital may lack credibility as counterparty and creates credit risk for the opposite party in derivatives trading. Therefore, the likelihood of holding trading derivatives is expected to decrease with *Lev* (Kim and Koppenhaver 1992).
Gap	FR Y-9C	Absolute value of the difference between bank assets and liabilities maturing or repricing within one year scaled by total assets. Banks with greater maturity mismatch and, therefore, greater exposure to interest rate risk are more likely to use derivatives to hedge their exposures (Sinkey and Carter 2000; Gunther and Siems 2002).
Nim	FR Y-9C	Net interest income scaled by total assets. Banks with lower net interest margins are more likely to use derivatives, particularly for trading purposes, in an attempt to increase fee income by selling derivatives or by speculating in derivative markets to offset reduced spreads from traditional banking activities (Shyu and Reichert 2002).
Npl	FR Y-9C	Ratio of the sum of nonaccrual loans and accruing loans past due 90 days or more to total assets. Lower loan quality is associated with higher credit risk. As part of a joint-risk management program, banks exposed to significant levels of credit risk are more likely to use derivatives to hedge other exposures (Schrand and Unal 1998).
Liquidity	FR Y-9C	Sum of cash and due, federal funds sold and securities purchased to resell scaled by total assets. Banks with greater liquidity are in a better position to manage large derivative positions (Jordan 1995) and are more likely to engage in derivatives trading activities.
Incentive	ExecuComp	Ratio of option vega to delta. Risk management theory identifies risk-taking incentives of managers as a determinant of risk management (Stulz 1984; Smith and Stulz 1985). Following Core and Guay (2002), we measure manager incentives to increase risk (stock price) by vega (delta), and include *Incentive* to capture the effect of managerial risk-taking incentives on the decision to use derivatives (Tufano 1996; Rogers 2002).

(continued on next page)

Variable Name	Data Source	Definition and Explanation
Convexity	Bank Compustat	Expected percentage tax savings from a 5 percent reduction in the volatility of taxable income. In the presence of tax function convexity, income smoothing reduces expected tax liabilities (Smith and Stulz 1985). Hedging serves as a mechanism to smooth income over time. Therefore, a number of studies (e.g., Nance et al. 1993; Dolde 1995; Mian 1996; Tufano 1996; Géczy et al. 1997; Graham and Rogers 2002) control for tax convexity as a determinant of derivative usage for hedging purposes. We measure tax convexity following Graham and Smith (1999).
ComIndLoan	FR Y-9C	Ratio of commercial and industrial loans to total assets. Commercial and industrial loan customers use derivatives to hedge their risks, and banks' familiarity with potential derivative customers increase the likelihood and extent of their participation in derivatives market as a dealer (Kim and Koppenhaver 1992).

APPENDIX C
DERIVATIONS OF UNADJUSTED SPREAD AND TAX-ADJUSTED SPREAD

Unadjusted Spread

We compute unadjusted bond yield spreads as the difference between the actual yield to maturity on the bank bond at issuance, and the yield to maturity on a synthetic Treasury bond, r_s, that perfectly matches the bank bond in terms of coupon payment, maturity, duration, and convexity.

In order to obtain r_s, we first treat each payment made by a coupon-bearing bank bond as a zero-coupon bond. For each of these payments, we calculate the risk-free spot rate using a parametric yield curve specification suggested by Gurkaynak et al. (2006). This specification, initially proposed by Nelson and Siegel (1987) and extended by Gurkaynak et al. (2006), is as follows:

$$r_t(n) = \delta_0 + \delta_1 \frac{1 - e^{\left(\frac{-n}{\tau_1}\right)}}{\frac{n}{\tau_1}} + \delta_2 \left[\frac{1 - e^{\left(\frac{-n}{\tau_1}\right)}}{\frac{n}{\tau_1}} - e^{\left(\frac{-n}{\tau_1}\right)} \right] + \delta_3 \left[\frac{1 - e^{\left(\frac{-n}{\tau_2}\right)}}{\frac{n}{\tau_2}} - e^{\left(\frac{-n}{\tau_2}\right)} \right], \tag{A.1}$$

where $r_t(n)$ is the yield to maturity of an n-year zero-coupon risk-free (Treasury) bond at time t; n is number of years till maturity; τ_1 and τ_2 are parameters that determine the locations of the "humps" in the yield curve. The parameters of this specification from 1961 to 2006 are provided on a daily basis by Gurkaynak et al. (2006).[31] Thus, for a given date during the 1961–2006 period, yields can be computed for future bond payments, each of which is assumed to be a zero-coupon bond, at all maturities. Next, we discount each payment made by the bond at a rate derived from Equation (A.1), and sum up the discounted values to obtain the price of a Treasury bond, which perfectly matches the bank bond in terms of coupon payment, maturity, duration, and convexity. The Treasury bond price, P^{TR}, is computed as:

[31] The parameters are also available at http://www.federalreserve.gov/pubs/feds/2006.

$$P^{TR} = \frac{C}{(1 + r_1)^1} + \frac{C}{(1 + r_2)^2} + \ldots + \frac{C}{(1 + r_n)^n} + \frac{F}{(1 + r_n)^n}, \tag{A.2}$$

where r_n is the yield derived from Equation (A.1) for a zero-coupon bond that matures in n years; C is the coupon payment; and F is the face value of the bond. In order to obtain the appropriate risk-free spot rate, we then solve the following equation for r_s:

$$P^{TR} = \frac{C}{(1 + r_s)^1} + \frac{C}{(1 + r_s)^2} + \ldots + \frac{C}{(1 + r_s)^n} + \frac{F}{(1 + r_s)^n}. \tag{A.3}$$

Tax-Adjusted Spread (Spread less Tax Premium)

In order to obtain a measure of tax-adjusted spread, r_{slt}, we reduce total spread, r_s, by the amount of estimated tax premium, r_t. We estimate r_t as the difference between the portion of spread explained by expected losses from default and taxes, and the portion explained by default premium only. Re-arranging the equations in Elton et al. (2001) yields:

$$r_t = \ln\left[\frac{D}{D - T}\right], \tag{A.4}$$

where $D = (1 - P_{t+1}) + \frac{\phi P_{t+1}}{V_{t+1T} + C}$, and $T = \frac{[C(1 - P_{t+1}) - (1 - \phi)P_{t+1}]}{V_{t+1T} + C} t_s(1 - t_f)$, and r_t is the tax premium; t_s is the state tax rate; t_f is the federal tax rate; and other terms are as defined in Equations (A.2) and (A.3). Spread less tax premium, r_{slt}, then equals the difference between r_s and r_t. We assume an effective tax rate, $t_s(1 - t_f)$, of 4.2 percent based on the grid search procedure in Elton et al. (2001). We examine different levels of tax rates between 3 percent and 5 percent at increments of 0.2 percent.[32] An effective tax rate of 4.2 percent minimizes the squared mean error between derived and actual sample bond prices. This effective tax rate is consistent with pairs of federal and state tax rates such as 33 percent and 6.3 percent, respectively.

REFERENCES

Ahmed, A. S., E. Kilic, and G. Lobo. 2006. Does recognition versus disclosure matter? Evidence from value-relevance of banks' recognized and disclosed derivative financial instruments. *The Accounting Review* 81 (3): 567–588.

America's Community Bankers. 1997. *Comment Letter to FASB Regarding SFAS 133*. Letter No. 132. Norwalk, CT: FASB.

Anson, M. J. P. 1999. *Accounting and Tax Rules for Derivatives*. New Hope, PA: Frank J. Fabozzi Associates.

Bank of America. 2006. *A Survey of the Effects of SFAS 133 on Corporate Hedging Practice*. Charlotte, NC: Bank of America.

Barth, M. E., W. H. Beaver, and W. R. Landsman. 1996. Value-relevance of banks' fair value disclosures under SFAS No. 107. *The Accounting Review* 71 (4): 513–537.

Beier, R. 1993. Will the FASB clip the hedges? *Financial Executive* 9 (4): 40–43.

Beaver, W., P. Kettler, and M. Scholes. 1970. The association between market-determined and accounting-determined measures of risk. *The Accounting Review* 45 (4): 654–682.

Bhojraj, S., and B. Swaminathan. 2009. How does the corporate bond market value capital investments and accruals? *Review of Accounting Studies* 14 (1): 31–62.

Billett, M. T., J. A. Garfinkel, and E. S. O'Neal. 1998. The cost of market versus regulatory discipline in banking. *Journal of Financial Economics* 48 (3): 333–358.

[32] This range is determined based on Elton et al.'s (2001) finding of an effective tax rate of 4 percent for their sample.

Booth, J. R., R. L. Smith, and R. W. Stolz. 1984. Use of interest rate futures by financial institutions. *Journal of Bank Research* 15 (1): 15–20.

Canadian Bankers Association. 1997. *Comment Letter to FASB Regarding SFAS 133*. Letter No. 36. Norwalk, CT: FASB.

Carpenter, J. N. 1996. Current issues in accounting for derivatives. *Journal of Derivatives* 3 (3): 65–71.

Consumer Federation of America. 1997. *Comment Letter to FASB Regarding SFAS 133*. Letter No. 18. Norwalk, CT: FASB.

Core, J., and W. Guay. 2002. Estimating the value of employee stock option portfolios and their sensitivities to price and volatility. *Journal of Accounting Research* 40 (3): 613–630.

Covitz, D. M., D. Hancock, and M. L. Kwast. 2000. *Mandatory Subordinated Debt: Would Banks Face More Market Discipline?* Washington, D.C.: Board of Governors of the Federal Reserve System.

Dadalt, P., G. Gay, and J. Nam. 2002. Asymmetric information and corporate derivatives use. *Journal of Futures Markets* 22 (3): 241–267.

Dolde, W. 1995. Hedging, leverage and primitive risk. *Journal of Financial Engineering* 4 (2): 187–216.

Eccher, A., K. Ramesh, and S. R. Thiagarajan. 1996. Fair value disclosures of bank holding companies. *Journal of Accounting and Economics* 22 (1–3): 79–117.

Elton, E. J., M. J. Gruber, D. Agrawal, and C. Mann. 2001. Explaining the rate spread on corporate bonds. *The Journal of Finance* 56 (1): 247–277.

Financial Accounting Standards Board (FASB). 1981. *Foreign Currency Translation*. Statement of Financial Accounting Standards No. 52. Stamford, CT: FASB.

———. 1984. *Accounting for Futures Contracts*. Statement of Financial Accounting Standards No. 80. Stamford, CT: FASB.

———. 1994. *Disclosure about Derivative Financial Instruments and Fair Value of Financial Instruments*. Statement of Financial Accounting Standards No. 119. Stamford, CT: FASB.

———. 1998. *Accounting for Derivative Instruments and Hedging Activities*. Statement of Financial Accounting Standards No. 133. Stamford, CT: FASB.

———. 2000. *Accounting for Certain Derivative Instruments and Certain Hedging Activities—An Amendment of FASB Statement No. 133*. Statement of Financial Accounting Standards No. 138. Stamford, CT: FASB.

———. 2003. *Amendment of Statement No. 133 on Derivative Instruments and Hedging Activities*. Statement of Financial Accounting Standards No. 149. Stamford, CT: FASB.

Financial Executives Institute. 1997. *Comment Letter to FASB Regarding SFAS 133*. Letter No. 21. Norwalk, CT: FASB.

Finnerty, J. D., and D. Grant. 2002. Alternative approaches to testing hedge effectiveness under SFAS No. 133. *Accounting Horizons* 16 (2): 95–108.

First Chicago NBD Corporation. 1997. *Comment Letter to FASB Regarding SFAS 133*. Letter No. 104. Norwalk, CT: FASB.

Flannery, M. J., and C. M. James. 1984. The effect of interest rate changes on the common stock returns of financial institutions. *The Journal of Finance* 39 (4): 1141–1153.

———, and S. Sorescu. 1996. Evidence of bank market discipline in subordinated debenture yields: 1983–1991. *The Journal of Finance* 51 (4): 1347–1377.

Francis, J., C. Lennox, and Z. Wang. 2010. Selection models in accounting research. Working paper, University of Missouri–Columbia and Nanyang Technological University.

Froot, K. A., D. S. Scharfstein, and J. C. Stein. 1993. Risk management: Coordinating corporate investment and financing policies. *The Journal of Finance* 48 (5): 1629–1658.

Gastineau, G., D. Smith, and R. Todd. 2001. *Risk Management, Derivatives, and Financial Analysis under SFAS No. 133*. Charlottesville, VA: Blackwell Publishers.

Gay, G. D., and J. Nam. 1998. The underinvestment problem and corporate derivatives use. *Financial Management* 27 (4): 53–69.

Géczy, C., B. A. Minton, and C. Schrand. 1997. Why firms use currency derivatives. *The Journal of Finance* 52 (4): 1323–1354.

Graham, J. R., and D. A. Rogers. 2002. Do firms hedge in response to tax incentives? *The Journal of Finance* 57 (2): 815–839.

Graham, J. R., and C. W. Smith. 1999. Tax incentives to hedge. *The Journal of Finance* 54 (6): 2241–2262.

Gunther, J. W., and T. F. Siems. 2002. The likelihood and extent of banks' involvement with interest rate derivatives as end users. *Research in Finance* 19 (1): 125–142.

Gurkaynak, R. S., B. Sack, and J. H. Wright. 2006. *The U.S. Treasury Yield Curve: 1961 to the Present*. Finance and Economics Discussion Series, Divisions of Research and Statistics and Monetary Affairs. Washington, D.C.: Federal Reserve Board.

Hancock, D., and M. L. Kwast. 2001. Using subordinated debt to monitor bank holding companies: Is it feasible? *Journal of Financial Services Research* 20 (2–3): 147–187.

Hodder, L., P. Hopkins, and J. Wahlen. 2006. Risk-relevance of fair-value income measures for commercial banks. *The Accounting Review* 81 (2): 337–375.

Hull, J. 1993. *Options, Futures, and Other Derivative Securities*. Englewood Cliffs, NJ: Prentice Hall.

Jagtiani, J., G. Kaufman, and C. Lemieux. 2002. The effect of credit risk on bank and bank holding company bond yields: Evidence from the post-FDICIA period. *Journal of Financial Research* 25 (4): 559–575.

Jordan, J. 1995. Supervision of derivative instruments. *Journal of Financial Services Research* 9 (3–4): 433–444.

———. 2000. *Depositor Discipline at Failing Banks*. Boston, MA: Federal Reserve Bank of Boston.

Kanagaretnam, K., G. V. Krishnan, and G. J. Lobo. 2010. An empirical analysis of auditor independence in the banking industry. *The Accounting Review* 85 (6): 1–36.

Kilic, E., G. J. Lobo, T. Ranasinghe, and K. Sivaramakrishnan. (2010). Earnings management and value relevance consequences of SFAS 133: Evidence from bank holding companies. Working paper, University of Houston, Houston, TX.

Kim, S., and G. D. Koppenhaver. 1992. An empirical analysis of bank interest rate swaps. *Journal of Financial Services Research* 7 (1): 57–72.

Koppenhaver, G. D. 1990. An empirical analysis of bank hedging in futures markets. *Journal of Futures Markets* 10 (1): 1–12.

Lee, L. 1983. Generalized econometric models with selectivity. *Econometrica* 51 (2): 507–512.

Liu, C., S. Ryan, and H. Tan. 2004. How banks' value-at-risk disclosures predict their total and priced risk: Effects of bank technical sophistication and learning over time. *Review of Accounting Studies* 9 (2–3): 265–294.

Liu, C., and S. Ryan. 2006. Income smoothing over the business cycle: Changes in banks' coordinated management of provisions for loan losses and loan charge-offs from the pre-1990 bust to the 1990s boom. *The Accounting Review* 81 (2): 421–441.

Mayers, D., and C. W. Smith. 1982. On the corporate demand for insurance. *The Journal of Business* 55 (2): 281–296.

McAnally, M. L. 1996. Banks, risk, and FAS 105 disclosures. *Journal of Accounting, Auditing & Finance* 11 (3): 453–490.

Mian, S. 1996. Evidence on corporate hedging policy. *Journal of Financial and Quantitative Analysis* 31 (3): 419–439.

Montesi, C., and T. Lucas. 1996. Toward improved accounting for derivatives and hedging activities. *The CPA Journal* 66 (10): 20–26.

Morgan, D. P., and K. J. Stiroh. 2001. Market discipline of banks: The asset test. *Journal of Financial Services Research* 20 (2–3): 195–208.

Nance, D. R., C. W. Smith, and C. W. Smithson. 1993. On the determinants of corporate hedging. *The Journal of Finance* 48 (1): 267–284.

Nelson, K. 1996. Fair value accounting for commercial banks: An empirical analysis of SFAS No. 107. *The Accounting Review* 71 (2): 161–183.

Nelson, C. R., and A. F. Siegel. 1987. Parsimonious modeling of yield curves. *The Journal of Business* 60 (4): 473–489.

Office of the Comptroller of Currency. 2005. *Interagency Advisory on Accounting and Reporting for Commitments to Originate and Sell Mortgage Loans*. OCC 2005-18. Washington, D.C.: Government Printing Office.

Puwalski, A. C. 2003. *Derivatives Risk in Commercial Banking*. Washington, D.C.: Financial Analysis Section of the Division of Insurance and Research, Federal Deposit Insurance Company.

Rane, D. 1992. A hedge or not a hedge? EITF 91–4 tried to answer the question. *Journal of Corporate Accounting & Finance* 3 (3): 279–284.

Rogers, D. A. 2002. Does executive portfolio structure affect risk management? CEO risk-taking incentives and corporate derivatives usage. *Journal of Banking & Finance* 26 (2–3): 271–295.

Ryan, S. G. 2007. *Financial Instruments and Institutions: Accounting and Disclosure Rules*. New York, NY: John Wiley & Sons, Inc.

Schrand, C. M. 1997. The association between stock-price interest rate sensitivity and disclosures about derivative instruments. *The Accounting Review* 72 (1): 87–109.

———, and H. Unal. 1998. Hedging and coordinated risk management: Evidence from thrift conversions. *The Journal of Finance* 53 (3): 979–1013.

Sengupta, P. 1998. Corporate disclosure quality and the cost of debt. *The Accounting Review* 73 (4): 459–474.

Shackelford, D., and T. Shevlin. 2001. Empirical tax research in accounting. *Journal of Accounting and Economics* 31 (1–3): 321–387.

Shyu, Y., and A. K. Reichert. 2002. The determinants of derivative use by U.S. and foreign banks. *Research in Finance* 19 (1): 143–172.

Sinkey, J. F., and D. A. Carter. 2000. Evidence on the financial characteristics of banks that do and do not use derivatives. *The Quarterly Review of Economics and Finance* 40 (4): 431–449.

Smith, C., and R. Stulz. 1985. The determinants of firms' hedging policies. *Journal of Financial and Quantitative Analysis* 20 (4): 391–405.

Stulz, R. M. 1984. Optimal hedging policies. *Journal of Financial and Quantitative Analysis* 19 (2): 127–140.

Swad, S. M. 1995. Accounting and disclosures for derivatives. Presentation at the 1995 Twenty-Second Annual National Conference on Current SEC Developments, Office of the Chief Accountant, U.S. Securities and Exchange Commission, Washington, D.C.

Tufano, P. 1996. Who manages risk? An empirical examination of risk management practices in the gold mining industry. *The Journal of Finance* 51 (4): 1097–1137.

Venkatachalam, M. 1996. Value relevance of banks' derivative disclosures. *Journal of Accounting and Economics* 22 (1–3): 327–355.

Wahlen, J. 1994. The nature of information in commercial bank loan loss disclosures. *The Accounting Review* 69 (3): 455–478.

Wong, M. H. F. 2000. The association between SFAS No. 119 derivatives disclosures and the foreign exchange risk exposure of manufacturing firms. *Journal of Accounting Research* 38 (2): 387–417.

THE ACCOUNTING REVIEW
Vol. 86, No. 3
2011
pp. 805–824

American Accounting Association
DOI: 10.2308/accr.00000034

Supply Chains and Segment Profitability: How Input Pricing Creates a Latent Cross-Segment Subsidy

Anil Arya
Brian Mittendorf
The Ohio State University

ABSTRACT: Recent years have seen an increased emphasis on developing more precise accounting measures of market- and customer-level profitability. By their very nature, such segment profitability measures may unwittingly neglect complementarities. Such complementarities have been widely recognized on the demand side due to brand recognition, predatory pricing, or interdependent products. We develop a model showing that important supply-side complementarities can also be prevalent. In particular, when a firm relies on a self-interested supplier for inputs used across multiple segments, the wholesale price it pays depends on the average profitability of its segments. Taking such interaction between upstream pricing and the firm's downstream reach into account, the model shows that: (1) segment profit calculations can understate or overstate the value added by the segment depending on the segment's relative contribution margin, and (2) the firm sometimes benefits from devoting resources to less profitable segments and perhaps even from serving seemingly unprofitable markets and/or customers.

Keywords: *input pricing; performance measurement; segment profit; supply chains.*

I. INTRODUCTION

Accountants have been persistent in attempts to generate disaggregate data that provides meaningful measures of segment profitability at both the market and customer levels. While joint costs and resource utilization have long been sticking points in such endeavors, recent developments stemming from activity-based costing (ABC) and customer lifetime value (CLV) analysis have been highly touted as means of better determining which markets and customers to service and the level of resources to devote to them. As a result, reliance on such measures has grown substantially in recent years.

We thank David Aboody, Mark Bagnoli, Anne Beyer, John Christensen, Peter Christensen, Sunil Dutta, Harry Evans (editor), John Fellingham, Hans Frimor, John Hughes, Steve Kachelmeier (senior editor), James Peck, Thomas Pfeiffer, Suresh Radhakrishnan, Madhav Rajan, Florin Sabac, David Sappington, Susan Watts, Lixin Ye, seminar participants at the University of Alberta, University of California, Los Angeles, The Ohio State University, Purdue University, The University of Texas at Dallas, University of Toronto, University of Vienna, and two anonymous referees for helpful comments. Professor Arya gratefully acknowledges support from the John J. Gerlach Chair.

Editor's note: Accepted by John H. (Harry) Evans III.

Submitted: December 2009
Accepted: October 2010
Published Online: April 2011

While the desire to finely parse performance is understandable, history is rife with examples in which refined performance measures failed to fully reflect opportunity costs of decisions.[1] In the context of market- and customer-level profitability, such opportunity costs have been recognized in terms of demand-side interdependencies. As examples, the benefits of having a loss-leader product for complementary demand-side effects on other products (e.g., "freebie marketing") or incurring sustained losses for competitive posturing (e.g., dumping or predatory pricing) have been well recognized. As such, practitioners and academics alike are acutely aware of demand-side considerations when making resource allocation choices.

In this study, we demonstrate that, in addition to the well-documented demand-side spillovers, supply-side spillovers can also be present, necessitating adjustments to segment profits in order to evaluate the value added by a segment. In particular, even if market and customer profitability measures closely track individual segment profitability, they do not fully reflect the cross-subsidization that can arise when a multi-segment firm relies on a supplier for key inputs. The cross-subsidization stems from the fact that supplier pricing is influenced by the extent and profitability of the buying firm's overall downstream reach.

To elaborate, the input price secured by a conglomerate depends on the conditions faced by the portfolio of markets in which it operates. As a result, altering the portfolio of markets that a firm serves alters the supply environment as well. A large retail conglomerate that has expanded beyond the "low-hanging fruit" of its prime markets to markets with more modest profitability can secure lower input prices. Conventional wisdom suggests such lower prices arise due to the buyer power that greater size affords. The results here suggest that there is more to buyer power than mere size; when a firm procures inputs for multiple uses, input prices are affected by average market conditions, and thus expansion to less profitable markets has the concomitant silver lining of lowering the prevailing input price.[2]

The basic model is of a firm that relies on a supplier for an input of material or labor that is used across multiple markets reflecting different segments or customers. The markets can differ in both the extent of retail demand and the degree of competition. With a presence in multiple markets, the seller is charged a wholesale price for the inputs it procures. Because the price does not dictate how or where the item has to be used, the firm is able to effectively prevent the input supplier from price discriminating based on usage. As such, when evaluating a segment's worth to the firm, the critical issue is not only an individual market's profitability, but also that market's effect on the firm's purchase sensitivity to input pricing.

Having identified this supply-side consideration, we next examine the circumstances under which such forces arise. In particular, given that the segments differ in both competition and retail demand, we identify the underlying drivers of cross-segment complementarities due to their effect on input price sensitivity. Interestingly, determining a segment's net effect on the firm's input price, and hence the direction of its supply-side effect, amounts to a concise comparison of a competition-weighted contribution margin for that segment to the competition-weighted average contribution margin of the other segments.

In this realm, practitioners routinely criticize firms for empire building and apparent corporate socialism, in that a firm that is highly successful in one arena tends to expand into other arenas

---

[1] Commonly discussed examples include the fact that accounting accruals occur in undiscounted terms and thus fail to reflect the cost of capital, product line costs do not reflect constrained capacity, and fixed cost allocations potentially induce a "death spiral" (see, for example, Zimmerman 2003).

[2] Consistent with this view, in recent times, major American auto makers have argued that their large dealer networks to rural and less-profitable metropolitan areas (as compared to the small, targeted dealership network of their foreign counterparts) make them deserving of lower prices from parts suppliers, more favorable labor contracts, and even cheaper capital (from government loans).

with less success. As an example, Toyota was widely panned for its push of the Lexus brand in Japan where the luxury vehicle market is dominated by European brands (e.g., Chozick 2007). Even if this effort does not yield the demand-side benefits of boosted sales in its home market, it can nonetheless help provide downward pressure on input prices that also affect its core business of luxury car sales in the U.S. In a similar vein, the slow start in U.S. sales of the new Acura ZDX does not bode as poorly for the firm as one may think, since the ZDX relies on as many common parts as the more widely profitable Acura MDX and Honda Pilot models and may, thus, help keep input prices low on these more profitable models.

Put differently, if the firm chooses to eliminate an underperforming segment and operate only its most successful product line, firm-wide profit would decrease by more-than-indicated segment profit figures alone due to the fact that the low-profit segment plays a large role in restricting the supplier wholesale price markup. Because such wholesale price adjustments are off-equilibrium, they are not reflected in (on-equilibrium) profit calculations and, as such, represent a latent subsidy provided by less-profitable markets to more-profitable markets.

As we demonstrate, this force means that it can even be the case that expanding into a seemingly unprofitable market may be optimal for a profit-maximizing firm. That is, when expanding (1) yields modestly negative profit in the new segment and (2) such losses are partly due to low margins of the new segment, it can be the case that expansion is worthwhile due to supply-side complementarities despite the negative segment profit that ensues. Our analysis provides precise conditions under which this phenomenon can (and cannot) occur. The theme of the analysis is further explored by examining the consequences of rivals' reliance on the same supplier as the multi-segment firm, and considering the effects of price competition in the retail realm. We demonstrate that this basic theme is robust with respect to these variations, though subtle new considerations arise.

The thrust of this study lies at the intersection of discussions about (1) measuring segment profitability and (2) coordinating self-interested supply chain partners. In the realm of profitability measurement, activity-based costing (ABC) has created a substantial improvement in the ability to track profits at the market or customer level (e.g., Cooper and Kaplan 1988). Through improved cost allocations, these processes are viewed by both academics and practitioners as critical to properly evaluating product line choices, resource allocation across segments, and keep-or-drop decisions (e.g., Goebel et al. 1998; Searcy 2004; Sopariwala 2005). At the customer level, ABC and customer lifetime value (CLV) calculations are used jointly to evaluate which customers to serve and target (e.g., Hogan et al. 2002; Searcy 2004; Venkatesan and Kumar 2004; Kuchta and Troska 2007). While significant efforts have been made to incorporate effects of downstream competition and demand-side complementarities in such calculations, our results point to strategic upstream supplier pricing as another issue worthy of consideration.

Supply chain frictions have been scrutinized by researchers in many areas of management. A fundamental concern at the forefront of such studies is the inherent conflict of interest in pricing. While supply chain partners have many common goals, individual interests can lead to inefficiencies that harm all. Perhaps most commonly analyzed among supply chain inefficiencies is that arising from double-marginalization. The issue of double-marginalization was first analyzed by Spengler (1950), following which several studies have considered the extent of pricing-induced distortions in supply chains and the myriad of means to alleviate such distortions (for summaries, see Katz (1989) and Lariviere [2008]).[3] We examine whether and how supply chain coordination concerns due to double marginalization interact with decisions a seller makes with regard to the

[3] Research in this area has also examined circumstances where prices above cost (and double-marginalization) prove helpful. Explanations include strategic customer behavior (Su and Zhang 2008), the time-inconsistency problem and

markets it serves. In doing so, we demonstrate a supply-chain-induced interdependency among product lines and customers in a conglomerate that is not reflected in standard segment profit calculations.

Extant research has also stressed other forms of supply chain inefficiency, manifest in terms of moral hazard among supply chain participants (e.g., Baiman et al. 2000; Balachandran and Radhakrishnan 2005; Hwang et al. 2006; Saouma 2008), adverse selection due to private information of supply chain members (e.g., Baiman and Rajan 2002), as well as a combination of these frictions (e.g., Baiman et al. 2001). In this vein, we stress that while our analysis examines a key interdependency among segments due to supply chain frictions in the context of double-marginalization, the key insight is not from double-marginalization *per se* but instead in showing that nontrivial supply chain frictions can induce supply-side spillovers that are not fully appreciated in segment profit calculations. Consistent with this point, we note that similar results can be obtained if the source of supply chain friction is due to adverse selection instead of double-marginalization.[4]

II. MODEL

A downstream (retail) firm relies on an upstream (wholesale) supplier for a key input. In converting the input and providing final goods to consumers, the downstream provider can reach multiple markets, some of which are more profitable than others. In particular, the retailer serves n segments, $n \geq 2$, and requires one unit of input per unit of output in each segment. In segment i, $i = 1,\ldots,n$, the retailer faces competition from a (Cournot) competitor, denoted rival i. For now, presume rivals do not rely on the inputs provided by the supplier. Retail demand in segment i for the firm (rival) is reflected by the inverse demand curve $p_i = a_i - q_i - \gamma_i\tilde{q}_i$ $(\tilde{p}_i = a_i - \tilde{q}_i - \gamma_i q_i)$, where p_i $(\tilde{p}_i)$ is the firm's (rival's) retail price for each unit, and q_i $(\tilde{q}_i)$ is the firm's (rival's) number of units sold. In the demand curve, a_i, $a_i > 0$, captures the degree of retail demand, and γ_i, $0 \leq \gamma_i \leq 1$ captures the degree of competitive intensity.[5]

Given these retail market conditions, the supplier determines the wholesale price for its inputs, w. For simplicity, we normalize the supplier's production cost and the firm's and rivals' other costs to zero. Upon observing the prevailing wholesale price, the firm and its rivals (concurrently) choose the quantity of output to provide to their respective retail markets. Denote the supplier's chosen wholesale price by w^n and the competitive retail outcome by $\boldsymbol{q}^n = \{q_i^n\}_{i=1}^n$ and $\tilde{\boldsymbol{q}}^n = \{\tilde{q}_i^n\}_{i=1}^n$ (in each case, the superscript n denotes the equilibrium when the firm participates in n markets). Given this outcome, segment i profit, $i = 1,\ldots,n$, is $\pi_i(q_i^n, \tilde{q}_i^n, w^n) = [a_i - q_i^n - \gamma_i\tilde{q}_i^n]q_i^n - w^n q_i^n$, and firm profit is $\Pi^n(\boldsymbol{q}^n, \tilde{\boldsymbol{q}}^n, w^n) = \Sigma_{i=1}^n \pi_i(q_i^n, \tilde{q}_i^n, w^n)$.

Given this setup, our interest is in determining the value added by a segment. Without loss of generality, consider the value of segment n. Using the above notation, in equilibrium, segment n's realized accounting profit is $\pi_n(q_n^n, \tilde{q}_n^n, w^n) \equiv \pi_n^*$. The value added by segment n is firm profit with segment n less firm profit that would be realized in the absence of segment n. That is, the value added by segment n, denoted V_n, is $V_n = \Pi^n(\boldsymbol{q}^n, \tilde{\boldsymbol{q}}^n, w^n) - \Pi^{n-1}(\boldsymbol{q}^{n-1}, \tilde{\boldsymbol{q}}^{n-1}, w^{n-1})$. The question we address in subsequent analyses is if and how accounting segment profit, π_n^*, differs from segment value added, V_n.

III. RESULTS

Equilibrium Supply Contracts

Before comparing the accounting measure of segment value with the underlying value added by a segment, we first derive the equilibrium outcome for operations in a given set of segments.

---

durable goods (Arya and Mittendorf 2006), and the desire to undercut *de facto* alliances between a supplier and a rival (Arya et al. 2008).

4 A previous version of this study provided details of such an analysis.

5 Throughout the analysis, we presume the parameters are such that there is nontrivial participation in each segment in equilibrium.

Then, comparing outcomes under n versus $n-1$ segments can provide a full characterization of the effects of segment n on firm profit. Working backward in the game, given the quoted wholesale price, w, and each rival's presumed retail quantity, $\tilde{q}_i$, the firm chooses retail quantities to maximize its total profit as in (1):

$$\underset{q_i, i=1,\ldots,n}{Max} \sum_{i=1}^{n} [a_i - q_i - \gamma_i \tilde{q}_i - w] q_i. \tag{1}$$

Similarly, each rival chooses $\tilde{q}_i$ to maximize its retail profit given q_i:

$$\underset{\tilde{q}_i}{Max} [a_i - \tilde{q}_i - \gamma_i q_i] \tilde{q}_i. \tag{2}$$

Jointly solving the first-order conditions of (1) and (2) yields:

$$q_i(w) = \frac{a_i[2-\gamma_i] - 2w}{4-\gamma_i^2} \quad \text{and} \quad \tilde{q}_i(w) = \frac{a_i[2-\gamma_i] + \gamma_i w}{4-\gamma_i^2}.$$

Intuitively; the higher the firm's marginal cost (w), the lower is its quantity and the greater is its rival's quantity. This feature gives rise to the supplier's induced demand, which governs the supplier's problem. In particular, the supplier chooses a wholesale price that maximizes its profit realized from the firm; this profit is derived from all n retail segments of the firm, as follows:

$$\underset{w}{Max} \sum_{i=1}^{n} w q_i(w) \Leftrightarrow \underset{w}{Max} \sum_{i=1}^{n} w \left[\frac{a_i[2-\gamma_i] - 2w}{4-\gamma_i^2} \right]. \tag{3}$$

The first-order condition of (3) is $\Sigma_{i=1}^{n} \frac{a_i[2-\gamma_i]-4w}{4-\gamma_i^2} = 0$. Denoting $k_i = \frac{1}{4-\gamma_i^2}$ and $\alpha_i = \frac{a_i}{2+\gamma_i}$ and the corresponding average values by $\bar{k}^n$ and $\bar{\alpha}^n$ (i.e., $\bar{k}^n = [1/n]\Sigma_{i=1}^{n} k_i$ and $\bar{\alpha}^n = [1/n]\Sigma_{i=1}^{n} \alpha_i$), the first-order condition yields $w^n = \bar{\alpha}^n / [4\bar{k}^n]$. Intuitively, the supplier's chosen wholesale price is affected by two key factors: the level of retail demand and the degree of competition. The greater the retail demand (the greater α_i), the greater is the firm's willingness to pay and, thus, the greater is the wholesale price. The greater the competitive intensity (the greater k_i), the greater is the threat that the rival will poach retail territory (and, hence, the supplier's wholesale territory) in response to a higher wholesale price. As a result, more competitive intensity translates into a lower chosen wholesale price. When this thinking is expanded to a firm serving n distinct retail markets, the supplier's price is similarly affected, but now the wholesale price reflects the cross-segment averages of each key factor ($\bar{\alpha}^n$ and $\bar{k}^n$).

Using this equilibrium wholesale price in the firm's and its rivals' quantity choices reveals equilibrium retail quantities, $q_i(w^n)$ and $\tilde{q}_i(w^n)$. Lemma 1 summarizes the outcome of this exercise. (All proofs are in the Appendix.)

Lemma 1: When the firm operates in n segments, the equilibrium outcome entails:

(i) $q_i^n = \frac{1}{2}\left[2\alpha_i - \frac{k_i}{\bar{k}^n}\bar{\alpha}^n\right]$, $\tilde{q}_i^n = \alpha_i + \frac{\gamma_i}{4}\frac{k_i}{\bar{k}^n}\bar{\alpha}^n$, and

(ii) $w^n = \frac{\bar{\alpha}^n}{4\bar{k}^n}$.

The lemma provides a crisp presentation of the above intuition. In Lemma 1(ii), as emphasized above, the supplier's wholesale price reflects a weighing of the retail demand and retail competition effects on wholesale demand. In each case, the relevant parameter reflects an average of those parameters that underlie each segment. The resulting procurement quantities reflect a

similar averaging. The quantity chosen by the firm in a given segment is increasing in that segment's demand (α_i) relative to the average demand of its segments ($\bar{\alpha}^n$). This latter effect means that the average demand of all segments, by boosting the wholesale price, serves to moderate the quantity choice of the firm in any given segment.

Similarly, the greater the competitive intensity a firm faces in any given segment (k_i) relative to its average intensity ($\bar{k}^n$), the lower is its quantity due to its rival grabbing market share. This consideration also has reverberations to the rivals' choices in that the firm's average market characteristics, by affecting wholesale prices and, in turn, the firm's retail choices, alter rival quantities. Thus, in terms of both a firm's competitive choices and its rivals' choices, a cross-segment interdependency arises solely due to upstream pricing considerations. In the case of the rival, this interdependency arises despite the fact that the rival does not rely on the supplier and participates in a single market. Given the equilibrium outcomes, we next consider if and how segment profit differs from segment value added.

Segment Value Added versus Segment Profit

Segment profit is the primary means through which a firm and outsiders view the contribution each segment makes to the firm's overall profit. When evaluating a segment's worth, it is often discussed that accounting profit measures can leave out positive or negative demand-side spillovers across segments. In particular, one product line may be a "loss leader" in that its segment profit is negative, yet it is worth maintaining due to its positive effect on the demand for other, more profitable, product lines. Inkjet printers represent a prominent example—printers are often sold near or below cost, but create a captive customer for cartridges that are sold at a substantial profit. In that case, looking at poor margins on printers reflected in segment profit clearly does not represent the whole story.

In a similar manner to such demand-side effects, this study seeks to identify the role of supply-side effects on segment value added. In both cases, accounting segment profit figures fairly reflect "what is," but may fail to answer "what if." Our measure of segment value added, on the other hand, is a measure of the incremental value of a segment, in that it compares current firm value with the value that would have been attained in the absence of that segment. As the following lemma demonstrates, the difference between segment profit and segment value added can be captured by supply-side and demand-side spillovers.

Lemma 2: Segment n's value added, V_n, can be expressed as:

$$V_n = \pi_n^* + S_n + D_n, \text{ where}$$

$$S_n = \Pi^{n-1}(\boldsymbol{q}^n, \tilde{\boldsymbol{q}}^n, w^n) - \Pi^{n-1}(\boldsymbol{q}^n, \tilde{\boldsymbol{q}}^n, w^{n-1}), \text{ and}$$

$$D_n = \Pi^{n-1}(\boldsymbol{q}^n, \tilde{\boldsymbol{q}}^n, w^{n-1}) - \Pi^{n-1}(\boldsymbol{q}^{n-1}, \tilde{\boldsymbol{q}}^{n-1}, w^{n-1}).$$

As Lemma 2 demonstrates, segment value added differs from segment profit due to supply-side spillovers (S_n) and demand-side spillovers (D_n). In particular, S_n reflects the change in profit of the n–1 remaining segments that would be induced by the concomitant change in supplier pricing, holding constant quantities demanded. Similarly, D_n reflects the change in profit of the n–1 remaining segments that would be induced by the concomitant change in retail quantities (demand), holding constant the wholesale price. While demand-side spillovers play a prominent role in the lexicon of business strategy as evidenced by the prominence in textbooks, academic research, and practitioner literature, the role of supply-side spillovers has not been demonstrated. The present analysis addresses this gap.

With this in mind, the most pertinent expression for the present analysis is S_n, which can be written more simply as $S_n = [w^{n-1} - w^n][q_1^n + \ldots + q_{n-1}^n]$. Thus, the sign of S_n amounts to determining whether segment n's presence serves to increase or decrease the prevailing wholesale price. If $S_n > 0$, then the wholesale price is reduced by operation in segment n; in such a case, segment n provides a latent supply-side subsidy to the other segments.

Determination of the extent of both the supply-side and demand-side effects requires substituting the equilibrium values $\boldsymbol{q}^n$, $\tilde{\boldsymbol{q}}^n$, $\boldsymbol{q}^{n-1}$, $\tilde{\boldsymbol{q}}^{n-1}$, w^n, and w^{n-1} from Lemma 1 into $\Pi^{n-1}(\boldsymbol{q}, \tilde{\boldsymbol{q}}, w)$. This exercise reveals:

$$S_n = \left[\frac{k_n}{4(n-1)\bar{k}^{n-1}}\right]\left[\frac{\bar{\alpha}^n}{\bar{k}^n} - \frac{\alpha_n}{k_n}\right]\left[\sum_{i=1}^{n-1} q_i^n\right], \text{ and}$$

$$D_n = -\left[\frac{k_n^2}{4(n-1)\bar{k}^{n-1}}\right]\left[\frac{\bar{\alpha}^n}{\bar{k}^n} - \frac{\alpha_n}{k_n}\right]\left[\left(\frac{\bar{\alpha}^n}{\bar{k}^n} - \frac{\alpha_n}{k_n}\right)\left(\frac{\sum_{i=1}^{n-1} k_i^2}{(n-1)\bar{k}^{n-1}}\right) - \frac{\sum_{i=1}^{n-1} k_i \gamma_i^2 q_i^n}{k_n}\right].$$

Summing S_n and D_n yields the following proposition:

Proposition 1: Segment n's profit understates the value it provides if it is weaker than average (i.e., $\frac{\alpha_n}{k_n} < \frac{\bar{\alpha}^n}{\bar{k}^n}$) and overstates value if it is stronger than average (i.e., $\frac{\alpha_n}{k_n} > \frac{\bar{\alpha}^n}{\bar{k}^n}$).

Two implications follow from the S_n and D_n expressions that underlie the proposition. First, the supply-side effect can either be positive or negative, depending on how the characteristics of the segment in question differ from that of other segments. Second, the demand-side effect is unequivocally negative. That is, this setting is one where the typical demand-side complementarity is not at play. Instead, if a complementarity is to arise, it is on the supply side. Importantly, this supply-side effect is of first-order importance, as evidenced by the fact that the condition that determines its sign is also the condition that determines the sign of the net effect, as confirmed in Proposition 1.

Proposition 1 completes the comparison of value added and segment profit. To be more precise, it confirms that $V_n - \pi_n^* = S_n + D_n = \lambda\left[\frac{\bar{\alpha}^n}{\bar{k}^n} - \frac{\alpha_n}{k_n}\right]$, $\lambda > 0$. In other words, despite the $2n$ parameters that underlie the setup ($\alpha_1 \ldots \alpha_n$ and $k_1 \ldots k_n$), the nature of the latent interdependency introduced by a segment can be succinctly captured by four parameters (α_n, $\bar{\alpha}^n$, k_n, and $\bar{k}^n$). To get a better intuitive feel for the condition in the proposition, we next isolate cross-segment variability in demand and competition, each in turn.

First, say the degree of competition is the same among all retail markets—i.e., $\gamma_1 = \ldots = \gamma_n$. In this case, with $k_n = \bar{k}^n$, the condition under which segment profit differs from value added amounts to a comparison of demand in that segment to the average demand among the portfolio of segments. In particular, segments with lower demand serve to indirectly subsidize those with higher demand due to their ability to moderate the prevailing wholesale price. This subsidy is captured by the segment's value added, but it is not reflected in the segment's profit number.

Corollary 1: In the $\gamma_1 = \ldots = \gamma_n$ case, segments with demand that is below (above) the average demand have profits that understate (overstate) the value they provide. That is, if $a_n < \bar{a}^n$, then $\pi_n^* < V_n$, and if $a_n > \bar{a}^n$, then $\pi_n^* > V_n$.

On the other hand, if demand is the same across segments—i.e., $a_1 = \ldots = a_n$, the only cross-market differences are reflected in varying degrees of competition. Note that the condition in the

proposition depends on α_n and $\bar{\alpha}^n$, which reflect competition-weighted measures of demand. So, despite having equal demand (intercepts), cross-market variability in the α_i terms can remain due to differences in competition. In this case, the condition in the proposition simplifies to a comparison of competition in segment n to the average competition level among the other segments. However, the average is not equally weighted; instead, the weights depend on the cross-segment variability in competition. This result is presented in Corollary 2.

Corollary 2: In the $a_1, \ldots, a_n$ case, segments with competition that is above (below) a weighted-average competition level have profits that understate (overstate) the value they provide. That is, if $\gamma_n > \Sigma_{i=1}^n\left[\frac{k_i}{n\bar{k}^n}\right]\gamma_i$, then $\pi_n^* < V_n$, and if $\gamma_n < \Sigma_{i=1}^n\left[\frac{k_i}{n\bar{k}^n}\right]\gamma_i$, then $\pi_n^* > V_n$.

The intuition for the results in Corollary 2 can be seen by examining the case of two segments. From the corollary, in the two-segment case with $a_1 = a_2$, profit in segment 2 understates value added if and only if $\gamma_2 > \gamma_1$. In effect, with $\gamma_2 > \gamma_1$, segment 2 provides a supply-side subsidy that is not picked up by accounting profit calculations, because competitive pressures in segment 2 help drive down the wholesale price that governs activity in both segments.

A natural follow-up question to the above analysis is how one could expect to see these parameters manifest in practice. While one may get a feel for the relative competitiveness of an industry or segment with ease, the notion of a higher or lower demand intercept is perhaps less natural. However, if one instead looks at profit margins, a related picture arises. In particular, denote the (unit) contribution margin of segment i by m_i, $m_i = a_i - q_i^n - \gamma_i\bar{q}_i^n - w^n$, and the n-segment average by $\bar{m}^n$, $\bar{m}^n = [1/n]\Sigma_{i=1}^n m_i$. Given this notation, the next proposition confirms that the succinct comparison in Proposition 1 also has an accounting analog.

Proposition 2: Segment n's profit understates the value it provides if its profit margin is weaker than average (i.e., $\frac{m_n}{k_n} < \frac{\bar{m}^n}{\bar{k}^n}$) and overstates value if its profit margin is stronger than average (i.e., $\frac{m_n}{k_n} > \frac{\bar{m}^n}{\bar{k}^n}$).

In short, the proposition shows that a segment that faces greater competition and/or lower contribution margins is one that serves an important off-equilibrium role in boosting the fortunes of other segments. Besides demonstrating the underlying forces driving an inherent inter-linkage in segments when a firm relies on external input supply, the proposition may shed some light on the widely held view that in multi-segment firms, underperforming segments are propped up by overperforming segments (e.g., Shin and Stulz 1998; Rajan et al. 2000; Gertner et al. 2002). Since such findings are derived in part from accounting segment profit measures, they may not reflect the latent input price subsidy provided in the reverse direction—i.e., a subsidy to overperforming segments provided by an underperforming segment.

Even more broadly, the proposition suggests that a firm with multiple segments that utilize common inputs would exhibit seemingly excessive (constrained) investment in underperforming (overperforming) segments, despite such behavior being value-maximizing. Viewed in this light, apparent inefficiencies in resource allocation may not be evidence of value destruction from diversification but rather reflect a form of selection bias in that multi-segment firms are inherently different from their single-segment counterparts (see, e.g., Chevalier 2004). Roughly stated, contrasting a multi-segment firm with a linear disaggregation of its segments misses out on the fact that third-party relationships of presumptive, stand-alone segments can be strikingly different from those nurtured by the aggregate firm. Clearly, such shifts must be accounted for in order to calculate a meaningful segment diversification discount/premium. If this view were borne out empirically, then one would expect the most (least) profitable divisions of a multi-segment firm to

perform better (worse) than their single-segment counterparts. Further, this feature would be expected to be more pronounced the more critical input supply is to the multi-segment firm.

Extending the theme, we next address if and how the latent supply-side subsidy provided by underperforming segments can alter the decision to enter a market in the first place.

The Decision to Enter a Market

The analysis thus far has demonstrated how accounting measures of segment performance can fail to capture the benefits of poorly performing segments due to their ability to put downward pressure on supplier pricing. Among other things, this view can possibly shed new light on seemingly widespread corporate socialism. When one thinks of criticisms of corporate socialism, wherein a firm uses one very successful line of business to seemingly support otherwise unsustainable segments, the notions of empire building and excessive market entry in response to success in a core market inevitably come to mind. To this end, we next consider the implications of the model for decisions to enter ancillary markets.

Suppose the firm operates in markets $1,\ldots,n-1$ and is considering entry into market n. Denoting the cost of entry into market i by F_i, the segment profit gleaned from entry into market n is simply $\pi_n^* - F_n$. Suppose further that this proposed expansion fits the bill for empire building in that the potential profit from the segment is below the average for the firm's core segments, i.e., $\frac{\alpha_n}{k_n} < \frac{\bar{\alpha}^n}{\bar{k}^n}$. In that case, from Proposition 1, entry into the new segment offers a latent benefit of favorably altering the prevailing wholesale price. In fact, this force, while latent, may form the basis for expansion into markets that yield negative profit. In other words, a firm may expand into new "loss-leader" markets, not as demand-side strategic posturing but instead as supply-side strategic posturing. The next proposition presents the conditions under which this phenomenon arises.

Proposition 3: If $\frac{\alpha_n}{k_n} < \frac{\bar{\alpha}^n}{\bar{k}^n}$, then the segment profit from entry is negative while the value added by entry is positive for all F_n in the non-empty interval $(F', F' + S_n + D_n)$, where $F' = \left[\alpha_n - \frac{k_n}{2\bar{k}^n}\bar{\alpha}^n\right]^2$.

In effect, the proposition notes that if a new segment is weaker than the existing portfolio of segments, then it is possible that entry therein yields negative segment profit but is nonetheless worthwhile. This occurs only for intermediate levels of entry costs. If the cost of entry is low ($F_n < F'$), then the positive margins from entry are sufficient to cover fixed costs, and segment profit in the new segment is positive. At the other extreme, for high entry costs ($F_n > F' + S_n + D_n$) the entry cost is so large that even when the supply-side complementarity is considered, entry is not worth the cost. For an intermediate interval of entry costs ($F_n \in (F', F' + S_n + D_n)$), the segment profit is negative but the supply-side subsidy makes entry worthwhile. Consistent with this, the larger the supply-side effect (S_n), the larger the interval.

Proposition 3 is couched in terms of implications of strategic input price adjustments on firms' entry decisions. Of course, in an analogous fashion, such concerns also impact a firm's decision to exit a market, where F_i represents the ongoing (fixed) cost of operating a segment. The example in Table 1 makes the point that the failure to recognize that the wholesale price is variable (not fixed) with respect to the decision to keep or drop a segment can lead to a death spiral. In a sense, this is the "reverse" of the familiar cost-allocation death spiral in which allocation can create an impression that a cost is variable while it really is fixed (and sunk). Broadly stated, both the supply-chain death spiral and the cost-allocation death spiral stress the need to be cognizant of the fixed versus variable cost dichotomy.

In Table 1, the first scenario characterizes the equilibrium outcomes when the firm operates in all three segments. In this case, the firm earns positive profit, but does so in the presence of a loss

TABLE 1

A Supply Chain Death Spiral

Example Parameters ($\gamma_i = 0$)

	Segment 1	Segment 2	Segment 3
Demand parameter (a_i)	40	30	20
Fixed cost (F_i)	110	45	10

Firm Operates in Segments 1, 2, and 3: (w = 15)

	Segment 1	Segment 2	Segment 3	Total
Output level (q_i)	12.50	7.50	2.50	22.50
Profits (π_i)	46.25	11.25	−3.75	53.75

Firm Operates in Segments 1 and 2: (w = 17.5)

	Segment 1	Segment 2	Total
Output level (q_i)	11.25	6.25	17.50
Profits (π_i)	16.5625	−5.9375	10.625

Scenario 3: Firm Operates in Segment 1: (w = 20)

	Segment 1	Total
Output level (q_i)	10	10
Profits (π_i)	−10	−10

in segment 3. The next scenario demonstrates the outcome of following a policy of exiting the unprofitable segment. In this case, the removal of segment 3 results in a lowering of firm-wide profit (from 53.75 to 10.625), in which case segment 2 arises as unprofitable. Eliminating this segment, in turn, leads to the final scenario, wherein only one segment remains and firm profit is negative. In other words, supplier pricing effects can introduce a death spiral of sorts, despite the fact that the fixed costs of operating in each segment are presumed to disappear along with the segment. Here, the death spiral arises because elimination of unprofitable segments serves to embolden the supplier in its pricing, thereby putting downward pressure on the margins of the more profitable segments. In effect, the firm's decision of eliminating a segment is not made in isolation in that the supplier, too, reacts strategically in response to such a course.

Rivals' Reliance on the Supplier

To this point, we have presumed a setting wherein a firm relies on the supplier for inputs but its rivals are self-sustaining. Suppose instead that the rivals, too, rely on the supplier for inputs. By providing a perspective on alternative supply-market conditions and the consequences for supply-side externalities from multi-market participation, this setting can serve as a robustness check on our underlying premise. In this case, the supplier sets $n+1$ wholesale prices, one for each firm (denoted w for the multi-segment firm and w_i for rival i). As before, the firm and its rivals concurrently choose retail quantities subsequent to observation of the prevailing wholesale prices.

Lemma 3 presents the equilibrium outcome under the setting with both the firm and its rivals reliant on the supplier, where the notation [′] distinguishes this equilibrium from that of Lemma 1.

Lemma 3: If the firm and the rivals each rely on the supplier, then the equilibrium outcome entails:

(i) $q_i'^n = \frac{1}{2}\left[\frac{a_i}{2+\gamma_i} - \frac{\bar{a}^n - a_i}{2}\right]$, $\tilde{q}_i'^n = \frac{a_i}{2[2+\gamma_i]}$, and

(ii) $w'^n = \frac{\bar{a}^n}{2}$, $\tilde{w}_i'^n = \frac{1}{2}\left[a_i + \frac{\gamma_i(\bar{a}^n - a_i)}{2}\right]$.

Comparing Lemmas 1 and 3, a key difference is apparent—whereas both the level of retail demand and the degree of retail competition played critical roles in the multi-segment firm's wholesale price in the former case, the wholesale price is independent of the degree of competition in the latter. That is, when the supplier provides inputs both to the firm and its rival, it does not tailor the firm's input price to the level of competition they face. To explain, when the supplier provides inputs only to the firm, it is that firm's *de facto* partner, and, as such, is eager to cut wholesale price to give the firm an edge in competition, i.e., as γ_i increases, w^n decreases. In contrast, when the rival is also the supplier's customer, the supplier has no incentive to favor one customer over another. The end result is that w'^n is independent of γ_i.

The desire to have a level playing field among competing firms extends to the wholesale price set for rival i, in that the rival's price reflects the average retail market conditions of the firm, the extent of which is dictated by the level of competition, γ_i. The effect of cross-segment averages extends to the rivals despite the fact that each participates in only one market. This feature arises so that if the firm benefits from an underperforming segment by getting price cuts, some of those cuts are carried over to the rival so as to maintain competition and thus boost input purchases in that market.

Given the induced wholesale prices and resulting competitive outcomes in the case of rival reliance on the supplier, the supply-side effect of participation in segment n is:

$$S_n' = \left[\frac{1}{2(n-1)}\right][\bar{a}^n - a_n]\left[\sum_{i=1}^{n-1} q_i'^n\right],$$

whereas the demand-side effect is:

$$D_n' = -\left[\frac{1}{16(n-1)}\right][\bar{a}^n - a_n]^2.$$

Given these expressions, the next proposition demonstrates the net effect of the firm's multi-market participation.

Proposition 4: With rival reliance on the supplier, segment n's profit understates the value it provides if its demand is weaker than average (i.e., $a_n < \bar{a}^n$) and overstates value if its demand is stronger than average (i.e., $a_n > \bar{a}^n$).

Comparing the expressions that underlie Proposition 4 and Proposition 1, it is readily seen that the rivals' reliance on the supplier simplifies the expression for both the demand- and supply-side effects. In particular, no longer is the degree of competition in any of the markets critical, but instead the distinguishing factor is the comparison of retail demand in one segment to the average retail demand levels. This factor carries forward to the determination of whether profit over- or understates segment value added. In particular, the difference between segment value added (V_n') and segment profit ($\pi_n'^*$) can be expressed as $V_n' - \pi_n'^* = S_n' + D_n' = \lambda'\lfloor\bar{a}^n - a_n\rfloor$, $\lambda' > 0$. This simplified expression means that a concise comparison of the demand level in segment n to the equally weighted average of that in other segments determines the sign of the difference between value added and profit.

Thus, besides demonstrating that the underlying premise herein is robust to who relies on the supplier for inputs, the setting also provides some guidance in terms of how the forces should arise in practice. That is, while retail demand in a segment relative to average demand is consistently a key determinant of any latent supply-side subsidies across segments, the degree of retail competition is crucially tied to the extent that the retail parties and the wholesale supplier form an alliance.

While the analysis so far has considered the cases wherein either the rivals are not reliant on a supplier (as in the base model) or reliant on a common supplier (as in this section), there are also supply arrangements in which each retail party relies on its own dedicated supplier. The following example contrasts the common supplier and the dedicated supplier settings, and reinforces the notion that supply-side effects can manifest themselves in a myriad of input procurement relationships.

Suppose there are two segments with $a_1 = a_2$. Moreover, for simplicity, say $\gamma_1 = 0$ and $\gamma_2 > 0$, i.e., the firm enjoys a monopoly in the first segment, while it faces a rival, rival 2, in the second segment. If the firm and the rival rely on a common supplier, then, from Lemma 3, the wholesale price is $a/2$ for each buyer. This is the same price that the supplier would have charged the firm had it operated only in segment 1. Thus, with a common supplier and $a_1 = a_2$, there are no supply-side effects and segment profit equals segment value added.

However, suppose the firm and the rival rely on separate (dedicated) suppliers. In this case, the wholesale prices for the firm and the rival are $\frac{a[2-\gamma_2][3\gamma_2+8]}{32-5\gamma_2^2}$ and $\frac{a[2-\gamma_2][16+4\gamma_2-\gamma_2^2]}{2[32-5\gamma_2^2]}$, respectively. Notice, the firm's input price is less than $a/2$, the price the firm would have been charged by its dedicated supplier had it operated only in segment 1. Thus, with dedicated suppliers, there is again a supply-side subsidy. Also, note that the wholesale prices for both parties are decreasing in γ_2. In effect, as segment 2 becomes more competitive, the suppliers respond by lowering the input charge to their respective clients. In fact, this reduced input-pricing effect can be so pronounced that both the firm and the rival may welcome the increased competition. In the example, the firm's profits are increasing in γ_2 for $\gamma_2 > 0.62$; the same is true for rival 2 for $\gamma_2 > 0.68$. The role of multi-segment operations is critical in deriving this result. After all, if the firm and the rival each operated only in segment 2, then each faces a wholesale price of $\frac{a[2-\gamma_2]}{4-\gamma_2}$. This price is again decreasing in γ_2 but each party would be made worse off by added competition. In contrast, the multi-segment firm gains because the increased competition in segment 2 lowers the input price that applies to purchases it makes in both segments, including the segment where it enjoys a monopoly and, thus, high profit margins. And, relatedly, when competition is sufficiently intense, rival 2 gains from the firm's ceding of market share in segment 2 in order to focus on its more profitable segment 1. Not surprisingly, then, similar results on retail parties benefiting from increased competition also apply when the multi-segment firm is the only party that relies on the supplier (as in the base model).

Price Competition

Thus far, we have considered the supply-side ramifications of a firm's participation in multiple output (retail) markets in a world of Cournot (quantity) competition. As the nature of competition has often been shown to lead to reversals of results (see, e.g., Darrough 1993; Dye 2001; Verrecchia 2001), we next consider the extent to which the key implications herein are sensitive to such variation by revisiting the analysis under Bertrand (price) competition. To reflect that the competitive outcome now arises under price competition, we denote all equilibrium outcomes in this case with a [$\hat{}$]. Following this notation, then, let $\hat{k}_i = \frac{2-\gamma_i^2}{[1-\gamma_i^2][4-\gamma_i^2]}$ and $\hat{\alpha}_i = \frac{a_i}{2+\gamma_i(1-\gamma_i)}$, and denote

the corresponding average values by $\bar{\hat{k}}^n$ and $\bar{\hat{\alpha}}^n$, respectively. Relegating the details to the Appendix, Lemma 4 presents the equilibrium outcomes in the case of price competition.

Lemma 4: Under retail price competition, the equilibrium outcome entails:

(i) $\hat{q}_i^n = \frac{1}{2}\left[2\hat{\alpha}_i - \frac{\hat{k}_i}{\bar{\hat{k}}^n}\bar{\hat{\alpha}}^n\right]$, $\hat{\bar{q}}_i^n = \hat{\alpha}_i + \frac{\gamma_i}{2[2-\gamma_i^2]}\frac{\hat{k}_i}{\bar{\hat{k}}^n}\bar{\hat{\alpha}}^n$, and

(ii) $\hat{w}^n = \bar{\hat{\alpha}}^n / [2\bar{\hat{k}}^n]$.

Using these outcomes, the price-competition analogs to the supply-side and demand-side effects from the main model are:

$$\hat{S}_n = \left[\frac{\hat{k}_n}{2(n-1)\bar{\hat{k}}^{n-1}}\right]\left[\frac{\bar{\hat{\alpha}}^n}{\bar{\hat{k}}^n} - \frac{\hat{\alpha}_n}{\hat{k}_n}\right]\left[\sum_{i=1}^{n-1}\hat{q}_i^n\right],$$

and

$$\hat{D}_n = -\left[\frac{\hat{k}_n^2}{2(n-1)\bar{\hat{k}}^{n-1}}\right]\left[\frac{\bar{\hat{\alpha}}^n}{\bar{\hat{k}}^n} - \frac{\hat{\alpha}_n}{\hat{k}_n}\right]\left[\left(\frac{\bar{\hat{\alpha}}^n}{\bar{\hat{k}}^n} - \frac{\hat{\alpha}_n}{\hat{k}_n}\right)\left(\frac{\sum_{i=1}^{n-1}(1-\gamma_i^2)\hat{k}_i^2}{2(n-1)\bar{\hat{k}}^{n-1}}\right) + \sum_{i=1}^{n-1}\frac{\gamma_i^2\hat{q}_i^n}{(4-\gamma_i^2)\hat{k}_n}\right],$$

respectively. Given these expressions, the analog to Proposition 1 is as follows:

Proposition 5: Under price competition, segment n's profit understates the value it provides if $\frac{\hat{\alpha}_n}{\hat{k}_n} < \frac{\bar{\hat{\alpha}}^n}{\bar{\hat{k}}^n}$ and overstates value if $\frac{\hat{\alpha}_n}{\hat{k}_n} > \frac{\bar{\hat{\alpha}}^n}{\bar{\hat{k}}^n}$.

In brief, comparing Propositions 1 and 5 confirms that the nature of competition (quantity versus price) does not alter the basic conclusions of the analysis. In fact, the basic form of both the supply- and demand-side effects of participation in segment n that are not accounted for in accounting segment metrics is invariant to the nature of competition. What changes instead is the precise calculation of the key parameters that capture retail demand (α_i versus $\hat{\alpha}_i$) and retail competition levels (k_i versus $\hat{k}_i$).

Taken together, the results throughout Section III suggest that (1) the importance and role of cross-segment supply-side subsidies induced by wholesale price effects is invariant to the nature of retail competition and the portfolio of supplier clients; and (2) the direction and magnitude of such supply-side effects can be critically affected by key characteristics of the retail markets (the degree of competition and extent of demand among the firm's portfolio of segments).

IV. CONCLUSION

Determining the contribution of various segments to the overall profit of a diversified entity is a delicate exercise. Refined costing measures have substantially improved this process and, hence, led to additional reliance on segment profit measures for resource allocation choices. Yet, because accounting profit metrics neglect opportunity costs, segment profit figures can deviate from the value added by a segment. The common concern in this vein is of downstream complementarities brought by loss-leader products or locations as means of furthering brand recognition and/or product loyalty. Our study examines a divergence between segment profit and segment value added that arises instead from upstream considerations.

We demonstrate that, even if product demand and long-term competitive posturing are not pressing considerations, a firm can nonetheless benefit from its weaker retail performers so as to

support more favorable supplier pricing terms. Since less profitable product lines or customer groups solidify a reluctance to accept markups in supplier pricing, a firm's active participation in less profitable markets can help support lower input prices. This effect, in turn, points to an additional consideration in keep-or-drop decisions.

While this fundamental phenomenon is demonstrated in a model wherein industrial structure is taken as given, further research could seek to endogenize the structure of industries and, in particular, the varied product offerings that a conglomerate would optimally seek in light of input-pricing considerations. Future study could also shed light on other forms of perceived "weakness" that could prove useful in supply chain relationships, such as public disclosure of proprietary information or even limited acquisition of relevant information.

APPENDIX

Proof of Lemma 1

Working backward, given wholesale price w and rival i's output $\tilde{q}_i$, the firm chooses quantity q_i in segment i, $i = 1,\ldots,n$, to maximize its total profit:

$$\underset{q_i, i=1,\ldots,n}{Max} \sum_{i=1}^{n} [a_i - q_i - \gamma_i \tilde{q}_i - w] q_i. \tag{A1}$$

Given q_i, rival i chooses $\tilde{q}_i$ to maximize its profit:

$$\underset{\tilde{q}_i}{Max} [a_i - \tilde{q}_i - \gamma_i q_i] \tilde{q}_i, \quad i = 1, \ldots, n. \tag{A2}$$

Jointly solving the first-order conditions of (A1) and (A2) yields:

$$q_i(w) = \frac{a_i[2-\gamma_i] - 2w}{4-\gamma_i^2} = \alpha_i - 2wk_i, \text{ and}$$

$$\tilde{q}_i(w) = \frac{a_i[2-\gamma_i] + \gamma_i w}{4-\gamma_i^2} = \alpha_i + w\gamma_i k_i, \quad i = 1, \ldots, n. \tag{A3}$$

The output reaction, $q_i(w)$ in (A3), serves as the induced demand function for the supplier, and it solves:

$$\underset{w}{Max} \sum_{i=1}^{n} w q_i(w) \Leftrightarrow \underset{w}{Max} \sum_{i=1}^{n} w[\alpha_i - 2wk_i]. \tag{A4}$$

The first-order condition of (A4) yields w^n in part (ii). Using (A3), the outputs listed in part (i) are $q_i(w^n)$ and $\tilde{q}_i(w^n)$. This completes the proof of Lemma 1. ■

Proof of Lemma 2

The value added by segment n is V_n, $V_n = \Pi^n(\boldsymbol{q}^n, \tilde{\boldsymbol{q}}^n, w^n) - \Pi^{n-1}(\boldsymbol{q}^{n-1}, \tilde{\boldsymbol{q}}^{n-1}, w^{n-1})$ Segment n's realized (accounting) profit is π_n^*, $\pi_n^* = \pi_n(q_n^n, \tilde{q}_n^n, w^n)$. Hence,

$$V_n - \pi_n^* = \Pi^n(\boldsymbol{q}^n, \tilde{\boldsymbol{q}}^n, w^n) - \Pi^{n-1}(\boldsymbol{q}^{n-1}, \tilde{\boldsymbol{q}}^{n-1}, w^{n-1}) - \pi_n(q_n^n, \tilde{q}_n^n, w^n). \tag{A5}$$

Since $\Pi^n(\boldsymbol{q}^n, \tilde{\boldsymbol{q}}^n, w^n) = \Pi^{n-1}(\boldsymbol{q}^n, \tilde{\boldsymbol{q}}^n, w^n) + \pi_n(q_n^n, \tilde{q}_n^n, w^n)$, (A5) can be written as:

$$V_n - \pi_n^* = \Pi^{n-1}(\boldsymbol{q}^n, \tilde{\boldsymbol{q}}^n, w^n) - \Pi^{n-1}(\boldsymbol{q}^{n-1}, \tilde{\boldsymbol{q}}^{n-1}, w^{n-1}). \tag{A6}$$

Adding and subtracting $\Pi^{n-1}(\boldsymbol{q}^n, \tilde{\boldsymbol{q}}^n, w^{n-1})$, (A6) can be written as:

$$V_n - \pi_n^* = [\Pi^{n-1}(\boldsymbol{q}^n, \tilde{\boldsymbol{q}}^n, w^n) - \Pi^{n-1}(\boldsymbol{q}^n, \tilde{\boldsymbol{q}}^n, w^{n-1})] + [\Pi^{n-1}(\boldsymbol{q}^n, \tilde{\boldsymbol{q}}^n, w^{n-1}) - \Pi^{n-1}(\boldsymbol{q}^{n-1}, \tilde{\boldsymbol{q}}^{n-1}, w^{n-1})]. \tag{A7}$$

The first term on the right-hand-side of (A7) is S_n and the second term is D_n. This completes the proof of Lemma 2. ■

Proof of Proposition 1

Because $\Pi^{n-1}(\boldsymbol{q}, \tilde{\boldsymbol{q}}, w) = \Sigma_{i=1}^{n-1} \pi_i(q_i, \tilde{q}_i, w)$ and $\pi_i(q_i, \tilde{q}_i, w) = [a_i - q_i - \gamma \tilde{q}_i - w] q_i$, the S_n and D_n expressions in Lemma 2 can be equivalently written as:

$$S_n = [w^{n-1} - w^n]\left[\sum_{i=1}^{n-1} q_i^n\right], \text{ and}$$

$$D_n = \sum_{i=1}^{n-1} \{[a_i - w^{n-1} - q_i^n - q_i^{n-1}][q_i^n - q_i^{n-1}] - \gamma_i[\tilde{q}_i^n q_i^n - \tilde{q}_i^{n-1} q_i^{n-1}]\}. \tag{A8}$$

From Lemma 1, $q_i^n = \frac{1}{2}\left[2\alpha_i - \frac{k_i}{\bar{k}^n}\bar{\alpha}^n\right]$, $q_i^{n-1} = \frac{1}{2}\left[2\alpha_i - \frac{k_i}{\bar{k}^{n-1}}\bar{\alpha}^{n-1}\right]$, $w^n = \frac{\bar{\alpha}^n}{4\bar{k}^n}$, $w^{n-1} = \frac{\bar{\alpha}^{n-1}}{4\bar{k}^{n-1}}$. Substituting these in (A8), and making use of the identities $\bar{\alpha}^n = \frac{n-1}{n}\bar{\alpha}^{n-1} + \frac{1}{n}\alpha_n$ and $\bar{k}^n = \frac{n-1}{n}\bar{k}^{n-1} + \frac{1}{n}k_n$, algebra yields the expressions for S_n and D_n provided in the text.

Using S_n and D_n:

$$V_n - \pi_n^* = \lambda\left[\frac{\bar{\alpha}^n}{\bar{k}^n} - \frac{\alpha_n}{k_n}\right], \text{ where}$$

$$\lambda = \left[\frac{k_n}{4(n-1)\bar{k}^{n-1}}\right]\left[\sum_{i=1}^{n-1} k_i \gamma_i^2 q_i^n + \sum_{i=1}^{n-1} q_i^n - \left(\frac{k_n \bar{\alpha}^n}{\bar{k}^n} - \alpha_n\right)\left(\frac{\sum_{i=1}^{n-1} k_i^2}{(n-1)\bar{k}^{n-1}}\right)\right]. \tag{A9}$$

The sign of $V_n - \pi_n^*$ is the same as the sign of $\frac{\bar{\alpha}^n}{\bar{k}^n} - \frac{\alpha_n}{k_n}$ if $\lambda > 0$. Note that $\lambda > 0$ if and only if the term $T = \left[\Sigma_{i=1}^{n-1} k_i \gamma_i^2 q_i^n + \Sigma_{i=1}^{n-1} q_i^n - \left(\frac{k_n \bar{\alpha}^n}{\bar{k}^n} - \alpha_n\right)\left(\frac{\Sigma_{i=1}^{n-1} k_i^2}{(n-1)\bar{k}^{n-1}}\right)\right] > 0$. Because $q_i^n > 0$, it follows that if $\alpha_n > \frac{k_n \bar{\alpha}^n}{\bar{k}^n}$, $T > 0$. Thus, the proof is complete if it can be shown that $T > 0$ even when $\alpha_n < \frac{k_n \bar{\alpha}^n}{\bar{k}^n}$. Given $\alpha_n < \frac{k_n \bar{\alpha}^n}{\bar{k}^n}$, $\alpha_n > \frac{k_n \bar{\alpha}^n}{2\bar{k}^n}$ (from $q_n^n > 0$ in Lemma 1), $1/4 \le k_i \le 1/3$ (with the limiting values corresponding to $\gamma_i = 0$ and $\gamma_i = 1$, respectively), a lower bound on $\Sigma_{i=1}^{n-1} q_i^n$ and an upper bound on $\left(\frac{k_n \bar{\alpha}^n}{\bar{k}^n} - \alpha_n\right)\left(\frac{\Sigma_{i=1}^{n-1} k_i^2}{(n-1)\bar{k}^{n-1}}\right)$ is derived below:

$$\sum_{i=1}^{n-1} q_i^n = \sum_{i=1}^{n} q_i^n - q_n^n = \frac{n\bar{\alpha}^n}{2} - \alpha_n + \frac{k_n \bar{\alpha}^n}{2\bar{k}^n} > \frac{n\bar{\alpha}^n}{2} - \frac{k_n \bar{\alpha}^n}{\bar{k}^n} + \frac{k_n \bar{\alpha}^n}{2\bar{k}^n} = \frac{n\bar{\alpha}^n}{2} - \frac{k_n \bar{\alpha}^n}{2\bar{k}^n} > \frac{n\bar{\alpha}^n}{2} - \frac{[1/3]\bar{\alpha}^n}{2[1/4]}$$
$$= \frac{n\bar{\alpha}^n}{2} - \frac{2\bar{\alpha}^n}{3}. \tag{A10}$$

$$\left(\frac{k_n\bar{\alpha}^n}{\bar{k}^n}-\alpha_n\right)\left(\frac{\sum_{i=1}^{n-1}k_i^2}{(n-1)\bar{k}^{n-1}}\right)<\left(\frac{k_n\bar{\alpha}^n}{\bar{k}^n}-\frac{k_n\bar{\alpha}^n}{2\bar{k}^n}\right)\left(\frac{[1/9][n-1]}{[1/4][n-1]}\right)=\frac{2k_n\bar{\alpha}^n}{9\bar{k}^n}<\frac{2[1/3]\bar{\alpha}^n}{9[1/4]}$$
$$=\frac{8\bar{\alpha}^n}{27}. \tag{A11}$$

From (A10) and (A11), and $n \geq 2$,

$$T>\sum_{i=1}^{n-1}k_i\gamma_i^2q_i^n+\frac{n\bar{\alpha}^n}{2}-\frac{2\bar{\alpha}^n}{3}-\frac{8\bar{\alpha}^n}{27}>\sum_{i=1}^{n-1}k_i\gamma_i^2q_i^n+\frac{\bar{\alpha}^n}{27}>0. \tag{A12}$$

Thus, $T > 0$ and, hence, $\lambda > 0$. This completes the proof of Proposition 1. ■

Proof of Corollary 1

In the $\gamma_1 = \ldots = \gamma_n$ case, denote γ_i by γ. Thus, $\frac{\alpha_n}{k_n}=a_n[2-\gamma]$ and $\frac{\bar{\alpha}^n}{\bar{k}^n}=\bar{a}_n[2-\gamma]$. Using these in (A9), $V_n-\pi_n^*=\lambda[2-\gamma][\bar{a}_n-a_n]$, $\lambda > 0$ and $0 \leq \gamma \leq 1$. This completes the proof. ■

Proof of Corollary 2

In the $a_1 = \ldots = a_n$ case, denote a_i by a. Thus, $\frac{\alpha_n}{k_n}=a[2-\gamma_n]$ and $\frac{\bar{\alpha}^n}{\bar{k}^n}=\frac{a}{n\bar{k}^n}\Sigma_{i=1}^n\frac{1}{2+\gamma_i}$. Using these in (A9), $V_n-\pi_n^*=\lambda a\left(\gamma_n-\Sigma_{i=1}^n\left[\frac{k_i}{n\bar{k}^n}\right]\gamma_i\right)$, $\lambda > 0$ and $a > 0$. This completes the proof. ■

Proof of Proposition 2

When the firm operates in n-segments, its (per-unit) contribution margin in segment i is $m_i = a_i - q_i^n - \gamma_i\bar{q}_i^n - w^n$. Using outputs and wholesale price in Lemma 1, and substituting a_i by $\alpha_i[2+\gamma_i]$ and γ_i^2 by $4 - 1/k_i$, m_i is:

$$m_i=\alpha_i-\frac{k_i}{2\bar{k}^n}\bar{\alpha}^n. \tag{A13}$$

From (A13), $\bar{m}^n=\frac{1}{n}\Sigma_{i=1}^n m_i=\frac{\bar{\alpha}^n}{2}$. Thus,

$$\frac{\bar{m}^n}{\bar{k}^n}=\frac{\bar{\alpha}^n}{2\bar{k}^n}\quad\text{and}\quad\frac{m_n}{k_n}=\frac{\alpha_n}{k_n}-\frac{\bar{\alpha}^n}{2\bar{k}^n}. \tag{A14}$$

From (A14), $\frac{\bar{m}^n}{\bar{k}^n}-\frac{m_n}{k_n}=\frac{\bar{\alpha}^n}{\bar{k}^n}-\frac{\alpha_n}{k_n}$. Thus, from (A9), $V_n-\pi_n^*=\lambda\left[\frac{\bar{m}^n}{\bar{k}^n}-\frac{m_n}{k_n}\right]$, $\lambda > 0$. This completes the proof of Proposition 2. ■

Proof of Proposition 3

Using Lemma 1 and (A13), segment n's profit from entry is:

$$\pi_n^*-F_n=m_nq_n^n-F_n=\left[\alpha_n-\frac{k_n}{2\bar{k}^n}\bar{\alpha}^n\right]^2-F_n=F'-F_n. \tag{A15}$$

From (A15), $F_n > F'$ ensures $\pi_n^*-F_n<0$. The value added by entry equals:

$$V_n-F_n=\pi_n^*+S_n+D_n-F_n=F'+S_n+D_n-F_n. \tag{A16}$$

From (A16), the firm prefers to enter if $F_n < F' + S_n + D_n$. Thus, for $F_n \in (F', F' + S_n + D_n)$, the segment profit from entry is negative while the value added by entry is positive. Finally, from

Lemma 2 and Proposition 1, the assumed condition $\alpha_n / k_n < \bar{\alpha}^n / \bar{k}^n$ implies $V_n - \pi_n^* = S_n + D_n > 0$. Thus, the interval $(F', F' + S_n + D_n)$ is non-empty. This completes the proof of Proposition 3. ■

Proof of Lemma 3

The firm's problem is again as in (A1). Given wholesale price $\tilde{w}_i$ and firm output q_i, rival i chooses $\tilde{q}_i$ to maximize its profit:

$$\underset{\tilde{q}_i}{Max}[a_i - \tilde{q}_i - \gamma_i q_i - \tilde{w}_i]\tilde{q}_i, \quad i = 1, \ldots, n. \tag{A17}$$

Jointly solving the first-order conditions of (A1) and (A17) yields:

$$q_i(w, \tilde{w}_i) = \frac{a_i[2 - \gamma_i] - 2w + \gamma_i \tilde{w}_i}{4 - \gamma_i^2}, \text{ and}$$

$$\tilde{q}_i(w, \tilde{w}_i) = \frac{a_i[2 - \gamma_i] - 2\tilde{w}_i + \gamma_i w}{4 - \gamma_i^2}, \quad i = 1, \ldots, n. \tag{A18}$$

The output reactions in (A18) serve as the induced demand function for the supplier, and it solves:

$$\underset{w, \tilde{w}_i, i=1,\ldots,n}{Max} \sum_{i=1}^{n} [w q_i(w, \tilde{w}_i) + \tilde{w}_i \tilde{q}_i(w, \tilde{w}_i)]. \tag{A19}$$

Solving the first-order conditions of (A19) yields w'^n and $\tilde{w}_i'^n$ in part (ii). Using these wholesale prices in (A18), the output decisions in part (i) equal $q_i(w'^n, \tilde{w}_i'^n)$ and $\tilde{q}_i(w'^n, \tilde{w}_i'^n)$. This completes the proof of Lemma 3. ■

Proof of Proposition 4

From Lemma 1, $q_i'^n = \frac{1}{2}\left[\frac{a_i}{2+\gamma_i} - \frac{\bar{a}^n - a_i}{2}\right]$, $q_i'^{n-1} = \frac{1}{2}\left[\frac{a_i}{2+\gamma_i} - \frac{\bar{a}^{n-1} - a_i}{2}\right]$, $w'^n = \frac{\bar{a}^n}{2}$, and $w'^{n-1} = \frac{\bar{a}^{n-1}}{2}$. Substituting these expressions in the analog to (A8), and making use of the identity $\bar{a}^n = \frac{n-1}{n}\bar{a}^{n-1} + \frac{1}{n}a_n$, algebra yields the expressions for S_n' and D_n'.

Using S_n' and D_n':

$$V_n' - \pi_n'^* = \lambda'[\bar{a}^n - a_n], \text{ where}$$

$$\lambda' = \left[\frac{1}{16(n-1)}\right]\left[8\sum_{i=1}^{n-1} q_i'^n + a_n - \bar{a}^n\right]. \tag{A20}$$

Using $q_i'^n$ from Lemma 3(i), the λ' expression in (A20) can be equivalently written as follows:

$$\lambda' = \left[\frac{1}{16(n-1)}\right]\left[8\sum_{i=1}^{n-1} \frac{a_i}{2(2+\gamma_i)} - \frac{8[(n-1)\bar{a}^n - (n\bar{a}^n - a_n)]}{4} + a_n - \bar{a}^n\right]$$

$$= \left[\frac{1}{16(n-1)}\right]\left[4\sum_{i=1}^{n-1} \frac{a_i}{2+\gamma_i} + \bar{a}^n - a_n\right]. \tag{A21}$$

If $a_n > \bar{a}^n$, from (A20), $\lambda' > 0$. And, if $a_n < \bar{a}^n$, from (A21), again $\lambda' > 0$. With $\lambda' > 0$, from (A20), the sign of $V_n' - \pi_n'^*$ is the same as the sign of $\bar{a}^n - a_n$. This completes the proof of Proposition 4. ■

Proof of Lemma 4

Solving $p_i = a_i - q_i - \gamma_i \tilde{q}_i$ and $\tilde{p}_i = a_i - \tilde{q}_i - \gamma_i q_i$, the demand functions are $q_i(p_i, \tilde{p}_i) = \frac{a_i[1-\gamma_i]-p_i+\gamma_i\tilde{p}_i}{1-\gamma_i^2}$ and $\tilde{q}_i(p_i, \tilde{p}_i) = \frac{a_i[1-\gamma_i]-\tilde{p}_i+\gamma_i p_i}{1-\gamma_i^2}$. Given wholesale price w and rival price $\tilde{p}_i$, in the Bertrand game, the firm chooses p_i in segment i, $i = 1,\ldots,n$, to maximize its total profit:

$$\underset{p_i, i=1,\ldots,n}{Max} \sum_{i=1}^{n} [p_i - w] q_i(p_i, \tilde{p}_i). \tag{A22}$$

Given retail price p_i, rival i chooses $\tilde{p}_i$ to maximize its profit:

$$\underset{\tilde{p}_i}{Max} \tilde{p}_i \tilde{q}_i(p_i, \tilde{p}_i), \quad i = 1, \ldots, n. \tag{A23}$$

Jointly solving the first-order conditions of (A22) and (A23) yields:

$$p_i(w) = \frac{a_i[2 - \gamma_i - \gamma_i^2] + 2w}{4 - \gamma_i^2}, \text{ and}$$

$$\tilde{p}_i(w) = \frac{a_i[2 - \gamma_i - \gamma_i^2] + \gamma_i w}{4 - \gamma_i^2}, \quad i = 1, \ldots, n. \tag{A24}$$

Using retail prices from (A24) in the demand functions $q_i(p_i, \tilde{p}_i)$ and $\tilde{q}_i(p_i, \tilde{p}_i)$ yields:

$$\hat{q}_i(w) = \frac{a_i[2 - \gamma_i - \gamma_i^2] - [2 - \gamma_i^2]w}{4 - 5\gamma_i^2 + \gamma_i^4} = \hat{\alpha}_i - w\hat{k}_i, \text{ and}$$

$$\hat{\tilde{q}}_i(w) = \frac{a_i[2 - \gamma_i - \gamma_i^2] + \gamma_i w}{4 - 5\gamma_i^2 + \gamma_i^4} = \hat{\alpha}_i + \frac{\gamma_i w \hat{k}_i}{2 - \gamma_i^2}, \quad i = 1, \ldots, n. \tag{A25}$$

(A25) serves as the induced demand function for the supplier, and it solves:

$$\underset{w}{Max} \sum_{i=1}^{n} w\hat{q}_i(w) \Leftrightarrow \underset{w}{Max} \sum_{i=1}^{n} w[\hat{\alpha}_i - w\hat{k}_i]. \tag{A26}$$

Solving the first-order condition of (A26) yields $\hat{w}^n$ in part (ii). Using this wholesale price in (A25), the equilibrium outputs, noted in part (i), equal $\hat{q}_i(\hat{w}^n)$ and $\hat{\tilde{q}}_i(\hat{w}^n)$. This completes the proof of Lemma 4. ■

Proof of Proposition 5

From Lemma 4, $\hat{q}_i^n = \frac{1}{2}\left[2\hat{\alpha}_i - \frac{\hat{k}_i}{\bar{\hat{k}}^n}\bar{\hat{\alpha}}^n\right]$, $\hat{q}_i^{n-1} = \frac{1}{2}\left[2\hat{\alpha}_i - \frac{\hat{k}_i}{\bar{\hat{k}}^{n-1}}\bar{\hat{\alpha}}^{n-1}\right]$, $\hat{w}^n = \frac{\bar{\hat{\alpha}}^n}{2\bar{\hat{k}}^n}$, and $\hat{w}^{n-1} = \frac{\bar{\hat{\alpha}}^{n-1}}{2\bar{\hat{k}}^{n-1}}$. Substituting these in the analog to (A8), and making use of the identities $\bar{\hat{\alpha}}^n = \frac{n-1}{n}\bar{\hat{\alpha}}^{n-1} + \frac{1}{n}\alpha_n$ and $\bar{\hat{k}}^n = \frac{n-1}{n}\bar{\hat{k}}^{n-1} + \frac{1}{n}k_n$, algebra yields the expressions for $\hat{S}_n$ and $\hat{D}_n$.

Using $\hat{S}_n$ and $\hat{D}_n$:

$$\hat{V}_n - \hat{\pi}_n^* = \hat{\lambda}\left[\frac{\bar{\hat{\alpha}}^n}{\bar{\hat{k}}^n} - \frac{\hat{\alpha}_n}{\hat{k}_n}\right], \text{ where}$$

$$\hat{\lambda} = \left[\frac{\hat{k}_n}{2(n-1)\bar{\hat{k}}^{n-1}}\right]\left[2\sum_{i=1}^{n-1}\frac{(2-\gamma_i^2)\hat{q}_i^n}{4-\gamma_i^2} - \left(\frac{\hat{k}_n\bar{\hat{\alpha}}^n}{\bar{\hat{k}}^n} - \hat{\alpha}_n\right)\left(\frac{\sum_{i=1}^{n-1}(1-\gamma_i^2)\hat{k}_i^2}{2(n-1)\bar{\hat{k}}^{n-1}}\right)\right]. \quad \text{(A27)}$$

The sign of $\hat{V}_n - \hat{\pi}_n^*$ is the same as the sign of $\frac{\bar{\hat{\alpha}}^n}{\bar{\hat{k}}^n} - \frac{\hat{\alpha}_n}{\hat{k}_n}$ if $\hat{\lambda} > 0$. Note that $\hat{\lambda} > 0$ if and only if the term $\hat{T} = 2\Sigma_{i=1}^{n-1}\frac{(2-\gamma_i^2)\hat{q}_i^n}{4-\gamma_i^2} - \left(\frac{\hat{k}_n\bar{\hat{\alpha}}^n}{\bar{\hat{k}}^n} - \hat{\alpha}_n\right)\left(\frac{\Sigma_{i=1}^{n-1}(1-\gamma_i^2)\hat{k}_i^2}{2(n-1)\bar{\hat{k}}^{n-1}}\right) > 0$.

Because $\hat{q}_i^n > 0$, it follows that if $\hat{\alpha}_n > \frac{\hat{k}_n\bar{\hat{\alpha}}^n}{\bar{\hat{k}}^n}$, $\hat{T} > 0$. Thus the proof is complete if it can be shown that $\hat{T} > 0$ even when $\hat{\alpha}_n < \frac{\hat{k}_n\bar{\hat{\alpha}}^n}{\bar{\hat{k}}^n}$. Given $\hat{\alpha}_n < \frac{\hat{k}_n\bar{\hat{\alpha}}^n}{\bar{\hat{k}}^n}$ and $1/3 \le \frac{2-\gamma_i^2}{4-\gamma_i^2} \le 1/2$ (with the limiting values corresponding to $\gamma_i = 0$ and $\gamma_i = 1$), a lower bound on $\hat{T}$ is derived below:

$$\hat{T} = 2\sum_{i=1}^{n-1}\frac{(2-\gamma_i^2)\hat{q}_i^n}{4-\gamma_i^2} - \left(\frac{\hat{k}_n\bar{\hat{\alpha}}^n}{\bar{\hat{k}}^n} - \hat{\alpha}_n\right)\left(\frac{\sum_{i=1}^{n-1}[(2-\gamma_i^2)/(4-\gamma_i^2)]\hat{k}_i}{2(n-1)\bar{\hat{k}}^{n-1}}\right) > \frac{2}{3}\sum_{i=1}^{n-1}\hat{q}_i^n - \left(\frac{\hat{k}_n\bar{\hat{\alpha}}^n}{\bar{\hat{k}}^n} - \hat{\alpha}_n\right)$$

$$\times\left(\frac{\sum_{i=1}^{n-1}\hat{k}_i}{4(n-1)\bar{\hat{k}}^{n-1}}\right) = \frac{2}{3}\sum_{i=1}^{n-1}\hat{q}_i^n - \frac{1}{4}\left(\frac{\hat{k}_n\bar{\hat{\alpha}}^n}{\bar{\hat{k}}^n} - \hat{\alpha}_n\right) = \frac{2}{3}\left(\frac{n\bar{\hat{\alpha}}^n}{2} - \hat{\alpha}_n + \frac{\hat{k}_n\bar{\hat{\alpha}}^n}{2\bar{\hat{k}}^n}\right) - \frac{1}{4}\left(\frac{\hat{k}_n\bar{\hat{\alpha}}^n}{\bar{\hat{k}}^n} - \hat{\alpha}_n\right)$$

$$= \frac{1}{3}(n\bar{\hat{\alpha}}^n - \hat{\alpha}_n) + \frac{1}{12}\left(\frac{\hat{k}_n\bar{\hat{\alpha}}^n}{\bar{\hat{k}}^n} - \hat{\alpha}_n\right) = \frac{1}{3}\left(\sum_{i=1}^{n-1}\hat{\alpha}_i\right) + \frac{1}{12}\left(\frac{\hat{k}_n\bar{\hat{\alpha}}^n}{\bar{\hat{k}}^n} - \hat{\alpha}_n\right). \quad \text{(A28)}$$

From (A28), for $\hat{\alpha}_n < \frac{\hat{k}_n\bar{\hat{\alpha}}^n}{\bar{\hat{k}}^n}$, $\hat{T} > 0$ and, hence, $\hat{\lambda} > 0$. This completes the proof of Proposition 5. ■

REFERENCES

Arya, A., and B. Mittendorf. 2006. Benefits of channel discord in the sale of durable goods. *Marketing Science* 25 (1): 91–96.

———, ———, and D. Sappington. 2008. The make-or-buy decision in the presence of a rival: Strategic outsourcing to a common supplier. *Management Science* 54 (10): 1747–1758.

Baiman, S., P. Fischer, and M. Rajan. 2000. Information, contracting, and quality costs. *Management Science* 46 (6): 776–789.

———, ———, and ———. 2001. Performance measurement and design in supply chains. *Management Science* 47 (1): 173–188.

———, and M. Rajan. 2002. The role of information and opportunism in the choice of buyer-supplier relationships. *Journal of Accounting Research* 40 (2): 247–278.

Balachandran, K., and S. Radhakrishnan. 2005. Quality implications of warranties in a supply chain. *Management Science* 51 (8): 1266–1277.

Chevalier, J. 2004. What do we know about cross-subsidization? Evidence from merging firms. *Advances in Economic Analysis & Policy* 4.

Chozick, A. 2007. The samurai sell: Lexus dealers bow to move swank cars; in Japan, Toyota revives ancient hospitality rites, but some buyers balk. *Wall Street Journal* (July 9): A1.

Cooper, R., and R. Kaplan. 1988. Measure costs right: Make the right decisions. *Harvard Business Review* 66 (September-October): 96–103.

Darrough, M. 1993. Disclosure policy and competition: Cournot vs. Bertrand. *The Accounting Review* 68 (3): 534–561.

Dye, R. 2001. An evaluation of "essays on disclosure" and the disclosure literature in accounting. *Journal of Accounting and Economics* 32 (1–3): 181–235.

Gertner, R., E. Powers, and D. Scharfstein. 2002. Learning about internal capital markets from corporate spin-offs. *The Journal of Finance* 57 (6): 2479–2506.

Goebel, D., G. Marshall, and W. Locander. 1998. Activity-based costing accounting for a market orientation. *Industrial Marketing Management* 27 (6): 497–510.

Hogan, J., D. Lehmann, M. Merino, R. Srivastava, J. Thomas, and P. Verhoef. 2002. Linking customer assets to financial performance. *Journal of Service Research* 5 (1): 26–38.

Hwang, I., S. Radhakrishnan, and L. Su. 2006. Vendor certification and appraisal: Implications for supplier quality. *Management Science* 52 (10): 1472–1482.

Katz, M. 1989. Vertical contractual relations. In *Handbook of Industrial Organization*, edited by R. Schmalensee, and R. Willig., Vol. 1. 655–721 London, U.K.: Elsevier.

Kuchta, D., and M. Troska. 2007. Activity-based costing and customer profitability. *Cost Management* 21 (3): 18–25.

Lariviere, M. 2008. Supply chains. In *The New Palgrave Dictionary of Economics*, edited by L. Blume, and S. Durlauf. 2nd edition. New York, NY: Palgrave Macmillan.

Rajan, R., H. Servaes, and L. Zingales. 2000. The cost of diversity: The diversification discount and inefficient investment. *The Journal of Finance* 55 (1): 35–80.

Saouma, R. 2008. Optimal second stage outsourcing. *Management Science* 54 (6): 1147–1159.

Searcy, D. 2004. Using activity-based costing to assess channel/customer profitability. *Management Accounting Quarterly* 5 (Winter): 51–60.

Shin, H., and R. Stulz. 1998. Are internal capital markets efficient? *The Quarterly Journal of Economics* 113 (2): 531–552.

Sopariwala, P. 2005. Evaluating the "keep-or-drop" decision. *Journal of Corporate Accounting & Finance* 16 (3): 47–61.

Spengler, J. 1950. Vertical integration and antitrust policy. *The Journal of Political Economy* 58 (4): 347–352.

Su, X., and F. Zhang. 2008. Strategic customer behavior, commitment, and supply chain performance. *Management Science* 54 (10): 1759–1773.

Venkatesan, R., and V. Kumar. 2004. A customer lifetime value framework for customer selection and resource allocation strategy. *Journal of Marketing* 68 (4): 106–125.

Verrecchia, R. 2001. Essays on disclosure. *Journal of Accounting and Economics* 32 (1–3): 97–180.

Zimmerman, J. 2003. *Accounting for Decision Making and Control*. New York, NY: McGraw-Hill Irwin.

THE ACCOUNTING REVIEW
Vol. 86, No. 3
2011
pp. 825–855

American Accounting Association
DOI: 10.2308/accr.00000036

Detection and Severity Classifications of Sarbanes-Oxley Section 404 Internal Control Deficiencies

Jean C. Bedard
Lynford Graham
Bentley University

ABSTRACT: We examine detection and severity classification of internal control deficiencies (ICD) identified under Section 404 of the Sarbanes-Oxley Act of 2002. While the cost/benefit balance of auditor testing of internal controls is highly controversial, prior research has not examined auditor versus client detection of ICD, nor has it examined factors auditors consider in judging ICD severity. We find that auditors detect about three-fourths of unremediated ICD, usually though control testing. This finding contrasts with extant research inferring control deficiency detection effectiveness from publicly available data, underscoring the value of Section 404 auditor testing in improving financial reporting quality. Auditors judge greater severity when a misstatement has already occurred. In the absence of a misstatement, severity is contingent on client and ICD characteristics, implying a more complex and nuanced judgment process without objective evidence of control failure. We also find that clients often underestimate ICD severity, but this tendency is lower among well-controlled companies with a well-designed Section 404 process.

Keywords: *internal controls; Sarbanes-Oxley Section 404; risk assessment; materiality.*

I. INTRODUCTION

This study investigates detection and severity classification of internal control deficiencies (ICD) under Section 404 of the Sarbanes-Oxley Act (SOX; U.S. Congress 2002). The purpose of Section 404 is to improve the reliability of information provided by public companies to the financial markets (COSO 2006; PCAOB 2004) by requiring company manage-

We thank the participating firms for their support of this research. This paper has benefited from comments of participants of the 2008 International Symposium on Auditing Research (Pasadena), the 2008 Conference of the Accounting and Finance Association of Australia and New Zealand (Sydney), the 2008 American Accounting Association Annual Meeting (Anaheim), and accounting research workshops at Bentley University, the College of William and Mary, and Northeastern University. We also appreciate helpful comments from Mohammad Abdolmohammadi, Carl Brewer, Christine Earley, Rani Hoitash, Udi Hoitash, Jim Irving, Karla Johnstone, Bill Kinney, Ganesh Krishnamoorthy, Mario Maletta, Tom Omer, Margie Platt, Tim Rupert, Roger Simnett, Stephen Taylor, Stuart Taylor, Jay Thibodeau, Kim Westermann, Arnie Wright, two anonymous reviewers, and Steve Kachelmeier (senior editor).

Editor's note: Accepted by Steven Kachelmeier.

Submitted: January 2009
Accepted: October 2010
Published Online: April 2011

ment and auditors to test internal control over financial reporting (ICFR) and to disclose severe control flaws that are not remediated as of the balance sheet date. Disclosure of material weakness(es) (MW) involves both detection and classification (Leone 2007), but in order to isolate these processes for study, information is needed on the base of ICD from which MW are derived.

Using proprietary data on detected ICD provided by several large audit firms, we address two research questions. First, what is the relative contribution of clients and auditors to ICD detection? This study is unique in its ability to examine, within the same time period, the extent of client (Section 404(a)) versus auditor (Section 404(b)) detection of ICD when both parties document and test controls. Within this analysis we examine factors contributing to client and auditor detection. One key feature of our study is that we provide direct evidence on the "yield" of detection methods used by auditors. This is an important issue at the heart of the debate on the value of auditor involvement in assessing and testing internal controls. Non-accelerated filers have recently been exempted by Congress from auditor attestation of controls effectiveness (Section 404(b)) as part of the financial reform legislation (U.S. Congress 2010). Emerging research (Kinney and Shepardson 2010; hereafter, KS) addresses the impact of auditor testing of controls using publicly available data. KS infer that because Audit Analytics (AA) data show that Section 404 reports for many companies with MW mention "material and/or numerous year-end adjustments," substantive testing might be sufficient to detect at least one MW for such companies, resulting in public reporting of ineffective controls. However, due to limitations of publicly available data, KS are not able to track misstatements to individual MW and, thus, their conclusion relates to company-level reporting of effective or ineffective controls. In contrast, we analyze deficiencies at the level of the individual control, directly examining detection method based on reports by engagement teams (as opposed to inferring possible detection method from annual public reports). Further, our data include not only MW, but also other detected ICD judged by auditors to be less severe.

Second, what factors are associated with the auditor's severity classifications of detected ICD? These classifications determine which control flaws are publicly reported (MW), and which are communicated to management and the audit committee (significant deficiency/deficiencies; SD).[1] In AS No. 2 (PCAOB 2004), SD are those deficiencies with a "more than remote" likelihood of failing to prevent/detect misstatements of "more than inconsequential" amount.[2] For MW, the amount of potential misstatement must be "material." These criteria for judging misstatement likelihood and materiality are likely complex to apply in practice due to difficulty in interpreting probability phrases, as recognized by prior research (Amer et al. 1994; Asare and Wright 2008). In addition, the wording of classification criteria is ambiguous, in implying that a deficiency can be "significant" (generally defined as important or consequential), but not be "material" (capable of influencing the judgments of financial report users). We study auditor judgments in the complex, unstructured task of severity classification, in this context of imprecise definitions and guidance.

We investigate these issues in a sample of audits of smaller accelerated filer companies, in which 3,990 ICD were detected and 2,942 remained unremediated at year-end.[3] Regarding our

---

1 Both classifications are consequential. Prior research shows that MW disclosure under Sections 404 or 302 is associated with adverse consequences to companies (e.g., Krishnan et al. 2008; Hammersley et al. 2008), and to company officers (Li et al. 2010). Because SD are not known unless voluntarily disclosed, research has not studied adverse consequences to company officers from SD detection. However, SD are also important because they may aggregate to a MW, or evolve into MW if not remediated in a timely manner (AS No. 2, ¶140).

2 AS No. 2, effective during our study period, was replaced by AS No. 5 (PCAOB 2007), which also requires that management's assertion and the auditor's opinion on ICFR effectiveness are supported by testing of controls. At inception in 2004, Section 404 applied to "accelerated filers," i.e., companies with $75 million of common equity float, among other criteria. Also during our study period, the SEC suggested that company managements look to PCAOB guidance for ICD classification. Specific SEC guidance for management was issued after our study period.

3 Severity classifications can only be studied for unremediated ICD. In order to examine detection and classification in the same sample, we restrict the detection analysis to unremediated ICD.

first research question, we find that clients detect fewer ICD than auditors, and are less likely to detect severe and pervasive ICD. Thus, many control flaws most likely to affect financial reporting would not be found in a client-driven process such as Section 302. Our results also show that use of a large accounting firm consultant for Section 404(a) work is associated with improved client detection. We extend this analysis by investigating how auditors detect ICD. We find that control tests provide initial evidence on a large proportion of ICD, including most MW and entity-level problems viewed as more serious by financial report users (Hammersley et al. 2008). This finding affirms auditors' Section 404 control testing as an important source of detecting control deficiencies, and implies that the recent exemption of Section 404(b) for smaller U.S. public companies could result in failure to fully realize potential improvements in financial reporting quality in that sector of the market.

Our second research question considers ICD severity classification. We first investigate the influence of a preliminary client classification (Earley et al. 2008). We find that clients tend to classify ICD as less severe, but auditors frequently override those classifications. Modeling ICD severity classifications, we find higher severity associated with: (1) greater knowledge and independence in the client's Section 404(a) process; (2) more objective evidence (e.g., an existing misstatement); (3) control flaws other than documentation problems (e.g., inappropriate design); (4) certain types of entity-level ICD (e.g., Control Environment); and (5) certain types of account-specific ICD (revenue and tax), consistent with the regulatory climate of the period.

Because prior research underscores the importance of objective evidence when standards are imprecise (e.g., Nelson et al. 2002), we further explore the role of an associated misstatements in ICD classification. An associated misstatement provides objective evidence that an ICD may fail again in the near future, but how do auditors judge severity when a misstatement has not already occurred? In this situation, we find that an independent client-reporting process, a longer auditor testing period, substantive test detection, and the number of MW identified become more important in classifying a given ICD as severe. Simply put, when objective evidence from a misstatement is not available, auditors use other evidence to support the classification decision.

In summary, our results support the value of auditor involvement at two stages of the ICFR assessment process (detection and classification), and contribute to understanding of factors associated with client and auditor performance in both stages. The following section provides background and contribution of this study, and develops our research questions and specific expectations. Section III discusses the study's methods, and Section IV contains results of models and other analyses. Section V presents the study's conclusions and limitations.

II. BACKGROUND, PRIOR RESEARCH, AND DEVELOPMENT OF RESEARCH QUESTIONS

Internal Control Regulation in the U.S. and Prior Related Research

SOX was implemented in response to numerous large-scale frauds and financial restatements among public companies. Among its key provisions is Section 404, which is based on the principle that improving ICFR should increase information reliability and restore investor confidence. This provision requires large public companies and their auditors to attest to ICFR effectiveness. Auditing standards prior to SOX included an assessment of internal controls as an element in obtaining the required understanding of the audited company and its environment.[4] Under those stan-

[4] An assessment of the control environment as a basis for risk assessment and planning is required in SAS No. 55 (AICPA 1988). SAS No. 78 (AICPA 1997) builds on that standard by incorporating the five components of the COSO Framework. Under AU 325, auditors were required to disclose observed "reportable conditions" (including material weaknesses in controls) to management and the audit committee or equivalent, but such communications were not made public. In 2003, the PCAOB adopted the AICPA Standards issued to date as interim standards.

dards, auditors had the option of relying on controls that were tested and found effective, or primarily using substantive procedures to support the financial statement audit opinion. Under prior standards, few auditors were familiar with documenting and testing of controls at the level expected by Section 404, using the COSO framework, and with the level of assurance required to issue a report on control effectiveness (see also KS; Curtis et al. 2009).

Prior research uses publicly available annual report data to distinguish characteristics of companies disclosing ineffective controls (i.e., at least one MW) under Section 404, or quarterly management reports under Section 302 (Ashbaugh-Skaife et al. 2007; Doyle et al. 2007; Hoitash et al. 2009). However, archival research does not address the full extent of detected control flaws, how those problems are detected, or how auditors determine which problems are disclosed. For example, KS use publicly available data on misstatements mentioned in Section 404 of the annual reports to infer that most companies would have disclosed ineffective controls (i.e., at least one MW detected) if a substantive approach were used. Because studies using publicly available data cannot directly examine detection method, however, research with access to control-level data is important. Further, studies using public reports have an incomplete data set from which to draw conclusions on detection method or the auditor's role, as only disclosure of unremediated MW is required. Evidence on auditors' severity classification decisions from behavioral experiments highlights the judgmental nature of these decisions. Earley et al. (2008) find that clients' preliminary severity classifications influence auditor classifications, which is especially problematic when clients propose low severity. Also, Wolfe et al. (2009) show that management persuasion tactics can affect auditors' classification of information technology controls.

In general, prior research implies that the component tasks of ICD severity classification (judging misstatement likelihood and materiality) are unstructured, complex, and difficult (e.g., Messier et al. 2005; Allen et al. 2006). One source of difficulty is interpretation of phrases such as "more than remote" and "material," derived from guidance on financial statement materiality (SAB No. 99; SEC 1999a) and accounting for contingencies (SFAS No. 5; FASB 1975). Another is inconsistency in definitions. In defining MW, the PCAOB follows the FASB's definition of "material" (SFAC No. 2, ¶132) as something that would change the judgment of a reasonable person relying on the information. Yet "significant" in the title of SD generally means something that is important, consequential, and large. Thus, auditing standards imply that an ICD can be important (thus a SD) but not affect users' decisions. Both issues lead to "ambiguity in the definitions" (Leone 2007, 227) of SD and MW, allowing latitude for company management in ICD classification. This ambiguity makes it difficult for auditors to judge the "right" ICD classification, and to justify that decision to client personnel whose incentive is not to report MW or SD. But to adhere to professional standards in a climate of liability concerns and PCAOB inspections, the auditor's incentive is to detect existing flaws in ICFR and bring them to the attention of client management, the audit committee, and the public if necessary. This study provides the first archival evidence of ICD detection and severity classification. In the following sections, we develop our research expectations from prior literature, auditing standards, and the economic context of the early years of Section 404 implementation.

Research Expectations

ICD Detection

We first focus on ICD detection, expecting that detection quality increases with expertise and independence of the parties performing the work. We expect that auditors will detect more ICD than clients, as they are independent from clients and have more knowledge of internal controls derived from education, firm training, and experience. We expect more detection among companies using Section 404 consultants, who are hired for their expertise and, in addition, have more independence from management than client personnel (although less than the external auditor). We

expect a greater impact when the consultant is from a large auditing firm due to significant training and guidance received in those firms. The relative knowledge of consultant personnel of other entities is less clear, as smaller firms likely did not institute the degree of training used by larger firms. In terms of relative knowledge/independence, we expect that companies with internal audit functions will outperform companies with no internal audit function.

In addition to these indicators of *who* is performing the client's control documentation and testing, we also consider the level of independence built into the client's process design. If company management is not involved in the reporting chain, then there is less opportunity for management influence. Thus, we expect a positive association of Section 404(a) process independence with client detection of ICD. We also consider the impact of integration of information technology (IT) in the Section 404(a) process, but do not predict a sign. While clients utilizing more advanced technology may detect more control flaws, those employing more IT may also have better underlying systems (i.e., fewer manual controls). Finally, we anticipate that a longer Section 404(a) testing period will result in better ICD detection.

We also investigate the nature of ICD detected by clients, an issue with important implications for understanding which control flaws are likely to be missed under the current regulatory structure for non-accelerated filers, which does not require auditor involvement. We expect that clients will be less likely to discover more severe problems, due to relative lack of knowledge about controls and the inherent objectivity challenge of "self-assessment." The issue of independence also arises in connection with client detection of entity-level ICD. Because many entity-level problems occur in centralized functions that directly involve top management, we expect that company personnel will be less likely to recognize flaws in controls at that level, relative to key Control Activities in areas such as sales and purchases.

ICD Severity Classification

Client characteristics and Section 404 processes. Our second summary research question concerns factors affecting auditors' severity classification decisions. We first consider client characteristics. Specifically, we expect higher ICD severity classifications in companies with lower overall quality of internal control.[5] In companies with good controls, an individual ICD is an isolated incident. However, in a weak-control context, auditors may consider a given ICD to be more severe, due to the potential for interaction with other control flaws and the potential for less effective compensating controls. We derive evidence on the overall controls context from the auditor's control reliance decision. We also consider the number of MW detected as an indicator of pervasive and serious problems in ICFR specifically (see also Hoitash et al. 2008; Hermanson and Ye 2009).

Our second severity expectation is that classifications will vary based on strength of the client's Section 404(a) processes. Consultants' greater expertise and independence could lead to better recognition of actual ICD severity, but these qualities could also result in remediation of more severe problems prior to year-end. Thus, we do not predict a directional effect for large firm or other Section 404 consultants. In addition, we expect that client Section 404(a) process independence will be positively associated with severity, as results of Section 404(a) activity will not be filtered through company management, which has the incentive not to report SD or MW.

We also consider the role of internal audit, proposing a nondirectional expectation. On the one hand, internal audit personnel may be more adept at recognizing ICD severity than other client personnel, if the internal audit unit has been involved in controls design. However, companies with

[5] Prior research (Ashbaugh-Skaife et al. 2007) tests the association of control context with MW disclosure indirectly, e.g., using size and financial position to proxy for investment in controls. We build on that study by isolating severity classification decisions and using more precise measures.

an internal audit could have less severe control flaws (Felix et al. 2001; Lin et al. 2011). We also consider the extent to which IT is integrated into Section 404(a) activity. Better use of IT could result in detection of more severe problems, as documentation and testing of controls will be more complete and accurate than a manual process. However, companies with greater use of IT for Section 404(a) likely also have better use of IT throughout their operations and, hence, may have better controls. Thus, we also do not predict a sign on this variable.

Auditor Section 404 processes. The auditor's Section 404(b) processes could also affect severity classification. First, we expect that severity classifications will be higher for auditor-detected ICD. Given their greater knowledge and objectivity, auditors should detect more serious control flaws than clients, and should be more confident of the severity implications of those they detect. The second factor in this category is the length of the auditor's testing process. During our sample period, anecdotal evidence from audit partners noted concern that clients' delay in completing their work could affect the quality of the auditor's process that follows. Further, due to constraints in the audit labor market during our study period, firms with heavy concentrations of accelerated filer clients struggled to hire and retain sufficient numbers of qualified employees to perform Section 404 procedures. Thus, an earlier start for auditor control testing should improve engagement teams' ability to fully consider and properly classify ICD.

Third, we consider the auditor's source of evidence. Specifically, we expect higher severity classifications associated with detection of ICD through substantive tests, and with an existing misstatement associated with the flawed control. Substantive tests are designed to detect account discrepancies that are the basis for proposed audit adjustments. If a monetary discrepancy has been detected, then evidence of ICD severity is stronger with respect to the likelihood criterion. Also, an actual misstatement amount creates a "floor" for the quantitative effect of the control deficiency, even though the potential severity could be greater than the amount observed. The presence of a misstatement is important, given the widespread use of quantitative criteria for determining materiality (e.g., Ng and Tan 2003; Messier et al. 2005). The association of misstatements with severity of control flaws is important to the argument proposed by KS that MW for many companies can be detected by tracing back from misstatements. If detection is through control tests and no associated misstatement has been detected, then the auditor's judgment in the Section 404 context becomes more complex—s/he must consider whether financial statement users' decisions could be affected by a *possible future* misstatement arising from a recognized ICD, without benefit of an actual current misstatement available as an example.

Prior research (Kennedy et al. 1997; Nelson et al. 2002; DeZoort et al. 2003) supports a positive association of misstatements with severity classification, by finding that objective evidence is important in justifying auditor decisions in unstructured tasks. These studies suggest that tangible evidence of control failure will improve auditors' ability to defend ICD classification decisions to management and audit committees, as well as in subsequent internal firm and PCAOB inspections. But how do auditors make severity classification decisions in the absence of a misstatement? We expect that in this case, stronger client and auditor processes will help bolster the auditor's position in discussions with management and the audit committee over ICD classification. Thus, process features such as an independent reporting chain and a longer auditor testing period could be more significant when a misstatement has not already occurred.

ICD type. We next consider whether severity classifications vary according to the nature of the ICD.[6] AS No. 2, ¶133 notes that the auditor should consider the nature of the affected financial

[6] Motivating study of differential classification according to nature of the control flaw, professional standards on Section 404 require specific consideration of accounts and assertions (AS No. 2, paras. 68–69). While past studies find little difference in control risk assessments across accounts/assertions (e.g., Elder and Allen 2003), current standards suggest that we could find systematic differences in severity classifications across types of ICD.

statement accounts in assessing the likelihood of misstatements that may result from an inoperative control. AS No. 2, ¶50–53 focus on entity-level controls with pervasive effects, including IT general controls (ITGC), controls over the financial reporting close process, and COSO categories of Control Environment, Monitoring, and company Risk Assessment. While some studies address the importance of Control Environment in control risk assessments (e.g., Cohen and Hanno 2000; Cohen et al. 2002), prior research has not directly examined how entity-level status affects auditors' severity classification in the Section 404 context.

Among account-specific ICD, we expect revenue accounts to be considered more severe. The SEC's strong focus on revenue recognition in the period immediately preceding the enactment of SOX (e.g., SAB No. 101; SEC 1999b) heightened auditors' awareness of the importance of controls over revenue recognition. We also predict a positive association of tax ICD with severity. In response to concerns about auditor independence, the SEC prohibited auditors from performing certain accounting services following passage of SOX, including preparation of the tax accrual. The transition away from auditor assistance to companies in tax accrual accounting to meet SEC/PCAOB independence requirements suggests higher likelihood of a material misstatement in tax accounts during our study period.

Professional guidance and some prior research on the second component judgment, materiality, also suggests differential classification according to the nature of the account. The definition of materiality in professional auditing standards is based on the auditor's judgment that an item would make a difference in users' decisions. AS No. 2, ¶23 notes that the auditor should consider both quantitative and qualitative elements, as in the financial statement materiality judgment (SAB No. 99; SEC 1999a). Qualitative factors noted in the standard include the nature of accounts and assertions, and future misstatement consequences. Prior research mainly considers materiality judgments in audit planning and audit adjustments. Summarizing this literature, Messier et al. (2005) conclude that during planning, the primary criterion for materiality in practice is income-based, although other criteria are considered. As noted above, judgments of materiality are likely more difficult when an existing misstatement is not available to provide a gauge for the likely amount of future misstatements resulting from the ICD.

In sum, this research builds on prior studies of publicly disclosed Section 404 MW by investigating detection and severity classifications of ICD discovered by companies and auditors. In the following section, we discuss our procedures and methods of testing expectations.

III. METHOD

Sample and Data Collection Procedures

We obtained data from several large audit firms, under confidentiality agreements that restrict us from knowing client identities and reporting separate results by firm or firm size. To guide sample selection, we asked that each firm randomly select from 2004–2005 engagements of smaller accelerated filers (with revenues of about $1 billion or less) in non-regulated industries. The restriction on client size is to improve our ability to generalize to the more than 10,000 smaller U.S. public companies not subject to Section 404(b) and to listed companies in other countries. Contact personnel from participating firms helped develop a spreadsheet to be completed by engagement teams, containing both company-level and control-level information.

Variable Definitions and Coding

Table 1, Panel A describes company-level variables representing auditors' and clients' Section 404 processes and outcomes, and Panel B describes control-level variables. We measure control quality using *IC_RELIANCE*, the auditor's level of reliance on internal controls. This is an indi-

TABLE 1

Client Controls Quality and Detection Process Measures

Panel A: Company-Level Client Control and Detection Process Variables: Mean (Std. Dev.) or Percent = 1

	All Engagements (n = 76)	Effective Controls (n = 61)	Ineffective Controls (n = 15)
IC_RELIANCE = 1 if current-year reliance on internal controls is "strong" or "maximum," and 0 otherwise.	48.7%	54.1%*	26.7%
CONSULTANT_BIG6 = percent of clients using a large audit firm as consultant for Section 404(a) activity.	13.2%	13.1%	13.3%
CONSULTANT_OTHER = percent of clients using a consultant other than a large audit firm.	63.2%	65.6%	53.3%
REPORT_INDEPENDENT = results of Section 404(a) activity are reported to the audit committee or auditor, independent of company management.	34.2%	32.8%	40.0%
INTERNAL_AUDIT = 1 if client has an internal audit function, and 0 otherwise.	46.1%	44.3%	53.3%
IT_INTEGRATION = auditor's judgment of the effectiveness of the client's integration of information technology into its 404(a) process (5 = highly effective to 1 = highly ineffective).	3.30 (0.88)	3.39 (0.88)	2.93 (0.80)**
MONTHS_PRIOR_CLIENT = months before year-end that client personnel began Section 404(a) testing.	5.82 (2.23)	5.91 (2.28)	5.47 (2.03)
MONTHS_PRIOR_AUDITOR = months before year-end auditor began Section 404(b) testing.	4.21 (1.76)	4.13 (1.70)	4.53 (2.03)
NUM_MW = number of material weaknesses.	2.01 (9.00)	0.00 (0.00)	10.20 (18.55)***
NUM_SD = number of significant deficiencies.	6.07 (14.06)	3.07 (4.88)	18.27 (28.61)***
NUM_DEF = number of deficiencies.	30.63 (41.76)	27.82 (38.35)	42.07 (53.51)

(continued on next page)

Panel B: Control-Level Variables (Percent = 1)

	Overall (n = 3,990)	Deficiencies (n = 2,328)	Significant Deficiencies (n = 461)	Material Weaknesses (n = 153)	Remediated (n = 1,027)
DETECTED_AUDITOR = identified by the auditor (or auditor and client jointly).	72.1%	74.7%	80.9%	84.3%	60.6%
DETECTED_SUBS = detected by substantive tests.	9.8%	9.2%	17.4%	26.1%	5.1%
DETECTED_CONTROL = detected by control tests.	63.5%	68.5%	67.5%	58.8%	56.1%
MISSTATEMENT_ASSOCIATED = ICD has failed to prevent/detect a misstatement, resulting in a current audit adjustment or restatement of prior periods.	11.2%	9.1%	20.0%	27.5%	9.6%
ENTITY = ICD at the entity level (COSO Control Environment, Monitoring, Information/Communications and Risk Assessment; General Ledger Controls, and Information Technology General Controls).	55.7%	55.6%	62.0%	46.1%	55.2%

*, **, *** Indicate ≤0.10, ≤0.05, and ≤0.01, respectively, with one-tailed probability levels for t-tests.

This table presents descriptive statistics on company-level and control-level variables.

Panel A presents company-level variables on client controls quality and detection processes of clients and auditors, with differences between engagements with effective/ineffective controls tested using independent sample t-tests or Pearson χ^2 tests (with d.f. = 1).

Panel B presents percentages of control-level variables, by severity category.

cator variable equaling 1 if "strong" or "maximum," and 0 otherwise.[7] *CONSULTANT_BIG6* equals 1 if the client outsourced all or part of its Section 404(a) activity to one of the largest six accounting firms, and 0 otherwise. *CONSULTANT_OTHER* equals 1 if the client outsourced its Section 404(a) activity to a smaller accounting firm or other provider. *REPORT_INDEPENDENT* equals 1 if Section 404(a) results are reported to the audit committee and auditor independent of client management, and 0 otherwise. *INTERNAL_AUDIT* equals 1 if the company has an internal audit function, and 0 otherwise. We capture the auditor's assessment of the company's effectiveness of integrating IT using a five-point scale (*IT_INTEGRATION*). *MONTHS_PRIOR_CLIENT* and *MONTHS_PRIOR_AUDITOR* measure the number of months before the balance sheet date that clients (auditors) began tests of controls. Table 1, Panel A also describes measures of the number of ICD at each severity level per engagement: *NUM_MW*, *NUM_SD*, and *NUM_DEF* for MW, SD, and deficiencies, respectively.

Control-level variables are described in Table 1, Panel B. Column headings indicate whether the ICD was remediated by year-end, and if not, the auditor's final severity classification. If the ICD was detected by the client, then the instrument asks for the client's preliminary severity classification, if known. *DETECTED_AUDITOR* equals 1 if detected by the auditor (or auditor and client jointly).[8] We also asked firms to indicate the phase of the audit in which the ICD was detected (planning, internal control assessment/testing, substantive testing, or review). *DETECTED_SUBS* equals 1 if the ICD was detected by substantive tests, and 0 otherwise; *DETECTED-_CONTROL* equals 1 if detected by substantive tests, and 0 otherwise. If the ICD was associated with a misstatement, then *MISSTATEMENT* equals 1, and 0 otherwise.[9] We developed additional coding rules from professional sources, verified by contact personnel at the firms providing data. We classified each ICD by the COSO components: Control Environment, Monitoring, Risk Assessment, Information/Communications, and Control Activities. We use AS No. 2 to determine which ICD are entity-level; i.e., those in the first four COSO categories (*ENTITY_CONTROLENVIR*, *ENTITY_MONITORING*, *ENTITY_INFOCOM*, and *ENTITY_RISKASSESS*), IT general controls (*ENTITY_ITGC*), and general ledger controls (*ENTITY_GL*). The indicator *ENTITY* equals 1 if the ICD is in any entity category, 0 otherwise.[10]

When applicable, ICD are classified by financial statement accounts/cycle, based on the COSO (2006) small business framework and guidance from some participating firms. These are: *ACC_PURCHASES*, *ACC_REVENUES*, *ACC_COMPENSATION*, *ACC_FIXEDINTANG* (fixed and intangible assets), *ACC_TAXES*, *ACC_LIABILITIES* (including leases, warranty and legal liabilities), and *ACC_TREASURY* (including investments and bank account management). We also classified each ICD by the nature of the control weakness, with categories derived from COSO and

---

7 Control reliance was assessed using the following scale: 5 = maximum reliance; 4 = significant reliance; 3 = some reliance; 2 = no reliance; 1= not applicable. Information on current-year reliance is missing for ten engagements. To enable use of this variable, we construct an indicator equaling 1 for maximum and significant reliance, and 0 otherwise. Measurement error from collapsing categories should bias against finding an effect of control reliance.

8 While auditing standards indicate that auditor testing should follow client testing within a given area, engagement teams indicated that for 37 ICD (1.3 percent), discovery was made by client and auditor jointly. Assuming that the clients might have missed these ICD without auditor involvement, we classify them with auditor-detected ICD.

9 Asking auditors to code associated misstatements was suggested by personnel of one of the participating firms, and so this question was asked of some but not all engagement teams. For remaining engagements, the authors identified associated misstatements from the text of the ICD description. If, for those engagements, the text did not mention a related misstatement when one existed, then the ICD would have been classified as not misstatement-associated. This should bias against finding differences between subsamples with/without associated misstatements.

10 We classify control flaws related to IT general controls (ITGC) based on AS No. 2 and guidance of the IT Governance Institute (2004) as system security, operations, development, and change process controls. General ledger issues relate to the following processes: financial statement close process, adjustments and accruals, allocations (e.g., depreciation and amortization), financial statement preparation, nonroutine entries, general ledger maintenance, SEC filings, and foreign currency.

AS No. 2: *DESIGN* (the control does not exist or cannot meet its objective); *DOCUMENTATION* (the control exists but is not documented; *INEFFECTIVE* (the control exists but does not function); *INSUFFICIENT_TESTING* (the control exists, but has not been adequately tested, perhaps due to insufficient time prior to year-end).[11]

Table 2 provides names and definitions for company-level control variables, based on prior research showing that MW disclosure is associated with company size, financial condition, complexity, and earnings quality (e.g., Ashbaugh-Skaife et al. 2007; Doyle et al. 2007).[12] We measure company size using the natural log of total assets (*LNASSETS*), and financial condition using total liabilities divided by total assets (*LEVERAGE*) and cash flow from operations (*CFOTA*). We also

TABLE 2

Sample Composition and Descriptive Statistics on Company-Level Control Variables Mean (Std. Dev.) or Percent = 1

	All Engagements (n = 76)	Effective Controls (n = 61)	Ineffective Controls (n = 15)
LNASSETS = natural log of assets.	19.52	19.54	19.43
	(1.16)	(0.95)	(1.21)
LEVERAGE = total liabilities/total assets.	0.43	0.44	0.40
	(0.22)	(0.22)	(0.17)
CFOTA = cash flow from operations/total assets.	−0.150	−0.012	−0.28
	(0.19)	(0.20)	(0.24)
ACCRUALS = residuals from a regression of current-year cash flows on net income, within industry categories.	0	−0.007	0.027
	(0.09)	(0.10)	(0.05)
ARINVTA = (accounts receivable + inventory)/total assets.	0.29	0.27	0.36
	(0.18)	(0.18)	(0.18)**
SEGMENTS = number of business segments.	2.51	2.38	3.07
	(2.32)	(2.01)	(3.35)
LOCATIONS = number of significant locations.	4.87	4.84	5.00
	(5.10)	(5.37)	(3.95)
NEW_PUBLIC = 1 if client has been a public company less than five years, and 0 otherwise.	30.3%	24.6%	53.3%**
INDUSTRY_MFG = 1 if client's industry is in the manufacturing sector (one-digit SIC = 2 or 3), and 0 otherwise.	56.6%	55.7%	60.0%
FYE_2004 = 1 if fiscal year is 2004, and 0 otherwise.	50.0%	47.5%	60.0%

** Indicates ≤0.05, significant effects in two-tailed tests.

This table presents descriptive statistics for company-level control variables, for the entire sample and for engagements with effective/ineffective ICFR. We test differences between companies with effective/ineffective ICFR using t-statistics for continuous variables and χ^2 tests (d.f. = 1) for dichotomous variables.

[11] To assure coding reliability, we used a practice sample of 73 ICD in two engagements, identified issues that reduced agreement, and refined our coding rules. We then independently coded 625 ICD from four engagements, with 97.2 percent agreement. The remaining ICD were coded by one or the other author.

[12] To protect client identities, financial information was disguised by engagement teams through multiplying by a factor between 0.95 and 1.05.

construct a measure of earnings quality by regressing net income on cash flow from operations within industry categories. The residuals from this regression constitute the variable *ACCRUALS*.[13] We measure complexity as the ratio of accounts receivable and inventory to total assets (*ARINVTA*), the number of reportable business segments (*SEGMENTS*), and the number of significant locations (*LOCATIONS*). We also include indicator variables for companies that have been public fewer than five years (*NEW_PUBLIC*), manufacturing companies (*INDUSTRY_MFG*),[14] and fiscal year using *FYE_2004*.

Models

We address our first research question in two ways: (1) by examining the overall percentage of ICD detected by clients/auditors; and (2) by modeling factors associated with likelihood of client detection. The latter analysis uses Model 1, a logistic regression model whose dependent variable is *DETECTED_CLIENT,* which equals 1 if the ICD was detected by the client, and 0 if detected by the auditor or auditor/client jointly.[15] Because we expect that clients are less likely to detect severe ICD, we expect negative coefficients on indicator variables *SD*, *MW*, and *ENTITY.* Because greater knowledge and independence from management of the parties performing the work should result in higher likelihood of client detection, we expect positive signs on *CONSULTANT_BIG6*, *CONSULTANT_OTHER*, *REPORT_INDEPENDENT*, and *INTERNAL_AUDIT*. Companies with better *IT_INTEGRATION* might perform better in ICD detection, but have better underlying controls, so we do not predict a sign on that variable. We expect a positive sign on *MONTHS_PRIOR_CLENT*, as a longer Section 404(a) process should help clients detect more ICD. We also expect a positive sign on *LNASSETS*, as larger companies should have more resources to devote to the process.

We investigate severity classification using two methods. First, because auditors are sometimes aware of the client's preliminary classification of ICD, we first test whether auditors override those classifications by judging ICD to be more severe. Second, we use two logistic regression models to test expectations regarding factors associated with severity classification of ICD, following the two-stage process implied in the definitions of SD and MW in AS No. 2 and explicitly featured in the profession's white paper (BDO Seidman LLP et al. 2004).[16] Model 2 (whose dependent variable is *SD/MW*; 1 = MW or SD, 0 = Deficiency), investigates factors associated with ICD meeting the criterion for a "more than remote" likelihood of failing to detect or prevent a misstatement (i.e., at least SD classification, implying that the ICD must be reported at least to management and the audit committee). Model 3 is tested on the subsample of ICD having met the likelihood criterion, investigating factors associated with the auditor's judgment that misstatements resulting from the ICD could be material (which determines public disclosure of ineffective ICFR). The dependent variable of Model 3 is *MW* (1 = MW, 0 = SD).

In Models 2 and 3, we expect more severe ICD among clients with lower control quality. Control quality is measured using *IC_RELIANCE* and the number of MW per engagement (due to

---

[13] Data limitations prevent us from estimating models to separate normal from abnormal accruals.

[14] Our sample comprises public entities in non-regulated industries. Most sample companies are in manufacturing (SIC codes ranging from 2000 to 3999; 56.6 percent). Because for 19 of our 76 engagements, engagement teams did not provide a SIC code, we construct an indicator variable comparing manufacturing to all other industries.

[15] Because there are multiple ICD for each company, all models are estimated with clustering on engagement ID (i.e., company/year), producing robust standard errors. Clustering is an important feature of our empirical tests, but it causes estimation difficulties when using some variables in smaller subsamples, due to insufficient numbers in one or both categories of the dependent variable. In that case, we omit as few variables as possible to enable estimation.

[16] Use of two models allows coefficients to vary across the two stages. To investigate whether this approach is statistically indicated, we estimated an ordered probit with a three-level dependent variable, and applied the Brant test for proportional odds. This test is significant ($p < 0.01$), indicating that the parallel regression assumption is violated and the two-model approach is appropriate.

skewness, we use *NUM_MW_SQRT,* the square root of the number of MW detected). We also investigate the association of ICD severity classification with client process measures.[17] We do not predict a sign on either *CONSULTANT* variable, as consultants may detect more severe control flaws, or enable earlier identification and remediation of severe flaws. We expect the sign on *REPORT_INDEPENDENT* to be positive, as it implies that Section 404 results are not filtered by management's incentives to classify ICD as less severe. We do not propose a sign on *INTERNAL_AUDIT*. Companies with an internal audit function might have less severe control flaws prior to Section 404 detection processes, and might also be better able to recognize severity than client personnel without internal audit experience. The sign on *IT_INTEGRATION* is also indeterminate. Companies with effective IT integration in the Section 404 process could perform better at detecting severe ICD. However, those companies likely also have better IT controls in general, meaning fewer severe problems should exist.

Models 2 and 3 include four auditor processes variables. We expect positive coefficients on *DETECTED_AUDITOR* (given auditors' greater knowledge and independence) and on *MONTHS_PRIOR_AUDITOR* (a longer auditor testing period provides more time to detect serious control flaws and consider their implications for financial reporting). We expect a positive sign on *MISSTATEMENT*, as a current misstatement provides objective evidence for assessing the likelihood and materiality of future misstatements. We also expect higher severity classifications for ICD detected with substantive tests (*DETECTED_SUBS*).

Regarding ICD type variables, we expect that entity-level ICD will be more severe, with one exception. Professional guidance during our study period (BDO Seidman LLP et al. 2004) indicates that regardless of the criteria of AS No. 2, auditors should classify ITGC consistent with underlying process controls. Thus, we do not propose a directional expectation on *ENTITY_ITGC*. Also, relative to the ICD in control documentation that remain in the intercept, we expect positive signs on *DESIGN*, *INEFFECTIVE*, and *INSUFFICIENT_TESTING*, as documentation problems should be more easily remediated than other types of control flaws.

We derive our predictions for company-level control variables from prior research on factors associated with public disclosure of MW (e.g., Doyle et al. 2007; Hoitash et al. 2009). We expect a negative sign on *LNASSETS*, as prior research shows smaller companies have more serious internal control problems. We expect that poor financial health will be associated with more severe ICD, generating a positive sign on *LEVERAGE* and a negative sign on cash flow from operations (*CFO*). We expect an association between ICD severity and *ACCRUALS*, based on prior research showing earnings quality to be lower among companies with internal control flaws (e.g., Ashbaugh-Skaife et al. 2009). We also expect that company complexity will be positively associated with ICD severity, including *ARINVTA*, *SEGMENTS*, and *LOCATIONS*. More severe ICD should be found among new public companies (*NEW_PUBLIC*). These models also control for industry membership using *INDUSTRY_MFG*, and for fiscal year using *FYE_2004*.

IV. RESULTS

Description of the Sample

The sample comprises 3,990 specific ICD detected in 76 engagements for 44 companies. An auditor's Section 404 opinion of ineffective ICFR was given for 15 sample engagements (19.7

[17] Companies' choices of Section 404 processes (e.g., to use a Section 404 consultant) could be endogenous. Prior research does not address factors associated with these choices; thus, there is little guidance in the literature on what measures could be used as instruments. Available variables within our sample are few, and none fit the criteria for effective instruments: a strong theoretical link, statistical association with the selection variable, and no association with the outcome variable (severity classification). In such a situation, Larcker and Rusticus (2010) advise against formal modeling of selection, as greater bias can be introduced by use of poor instruments.

percent).[18] Untabled analysis shows that an additional 51.4 percent of sample companies have at least one SD, implying that only 28.9 percent had no ICD of sufficient severity to be reported, at least to the audit committee. At the individual control level, 153 are MW (3.8 percent of detected ICD), a mean (median) of 10.2 (2.0) MW per company with ineffective controls. Most companies with ineffective controls have few MW, but some have a large number of ICD that auditors identified as MW or components of MW. However, this does not imply that each of these individual control flaws would be disclosed separately in Section 404 reports. Such disclosures vary in their level of specificity; for instance, individual deficiencies might be aggregated for discussion in the text of the opinion (e.g., several control flaws in revenues or several issues relating to the Control Environment). SD comprise 11.6 percent of sample ICD, and Deficiencies 58.4 percent, while 25.7 percent were remediated as of the balance sheet date. Details on ICD type by severity level are shown in Table 3.[19]

Table 2 presents descriptive statistics for company-level variables, and tests differences between subsamples with effective versus ineffective controls. The mean log of total assets is 19.52 (the mean in dollars is \$502.5 million), mean leverage is 0.43, and mean *CFOTA* is −0.150. The mean of *ACCRUALS* (the residual of regressing net income on cash flow from operations) is zero, by construction. The mean ratio of *ARINVTA* is 0.29; this ratio is larger for companies with ineffective ICFR (0.36 versus 0.27, $p = 0.048$). On average, sample companies have 2.51 segments and 4.87 locations. The overall percentage of companies that have been public fewer than five years is 30.3 percent; companies with ineffective controls are significantly more likely to be newly public (24.6 versus 53.3, $p = 0.030$). Over half of the companies (56.6 percent) are in the manufacturing industry.

Descriptive Statistics on ICD Detection

Table 1, Panel A describes company-level Section 404 process variables. Panel A shows that control reliance is higher in engagements with effective controls (54.1 versus 26.7 percent, $p = 0.057$). Consultants from large auditing firms were used for Section 404 activity by 13.2 percent of sample companies, while 63.2 percent used a smaller auditing firm or another provider.[20] Results of the client's Section 404 activity were reported independent of management in 34.2 percent of the sample. Thus, in about two-thirds of sample companies, Section 404 results were filtered through management before being made available to the auditor. Just under half of the companies (46.1 percent) have an internal audit unit. Companies with ineffective controls have a lower mean *IT_INTEGRATION* than those with effective controls (2.93 versus 3.39 on the scale of 1 = highly ineffective to 5 = highly effective, $p = 0.035$). Clients (auditors) began control tests an average of 5.82 (4.21) months before year-end.[21] Panel A also shows that the mean number of MW among engagements with ineffective controls is 10.20, while the median is 2.01, indicating a

[18] Bedard et al. (2007) note that the overall proportion of smaller accelerated filers with ineffective controls under Section 404 is 20 percent in 2004 and 14 percent in 2005. Our sample proportions of companies with ineffective ICFR in 2004 (2005) are 23.7 (15.8) percent, not significantly different from the Bedard et al. (2007) disclosure proportions ($p = 0.345$ and $p = 0.444$, respectively).

[19] For efficiency, we do not present a matrix of correlations among independent variables. Estimating the severity models using regression to obtain variance inflation factors, we find that the highest is 4.42 (4.41) for *NUM_MW_SQRT* in Model 2 (3), lower than the level of 5 often used to indicate concern for multicollinearity.

[20] Untabled statistics show that the consultant managed the work in 35.1 percent of companies using a consultant, performed controls documentation in 68.4 percent, and controls testing in 86 percent.

[21] There were numerous reports in the business press in the first year of Section 404 implementation of concern that some companies delayed their start of testing too long to do an effective job. In supplemental analysis, we investigate whether the client's Section 404 testing began earlier for 2005 engagements in our sample than for 2004 engagements. In 2004, client testing began 5.28 months before year-end, while in 2005 client testing began 6.34 months prior to year-end. This difference is significant ($t = 2.087$, $p = 0.020$).

TABLE 3

Percentage Composition of Detected ICD by Severity Category

Panel A: Entity-Level ICD

	All (n = 3,990)	Deficiencies (n = 2,328)	Significant Deficiencies (n = 461)	Material Weaknesses (n = 153)	Remediated (n = 1,027)
ENTITY_MONITORING = COSO category of Monitoring.	12.6%	12.0%	14.3%	11.2%	13.1%
ENTITY_CONTROLENVIR = COSO category of Control Environment.	7.2	8.4	6.3	9.2	4.8
ENTITY_INFOCOM = COSO category of Information/Communications.	1.7	1.7	0.9	2.0	2.4
ENTITY_RISKASSESS = COSO category of Risk Assessment.	0.3	0.4	0	0.7	0.1
ENTITY_GENERAL_LEDGER = controls over general ledger processes, including accruals, closing, consolidations, and SEC filings.	22.0	19.9	27.1	26.3	24.1
ENTITY_ITGC = information technology general controls.	21.4	21.9	26.3	4.6	20.9
Total entity-level controls	55.7%	55.6%	62.0%	46.1%	55.2%

Panel B: Type of Control Failure

	All (n = 3,990)	Deficiencies (n = 2,328)	Significant Deficiencies (n = 461)	Material Weaknesses (n = 153)	Remediated (n = 1,027)
INEFFECTIVE = ineffective control.	41.5%	39.3%	43.6%	54.6%	43.1%
DOCUMENTATION = failure to document a control.	28.6	30.7	21.5	17.8	29.0
DESIGN = control design flaw.	25.2	23.8	34.7	25.7	24.1
INSUFFICIENT_TESTING = client failure to sufficiently test a control.	8.7	8.8	9.5	16.5	6.9

Panel C: Cycle/Account

	All ICD (n = 3,990)	Deficiencies (n = 2,328)	Significant Deficiencies (n = 461)	Material Weaknesses (n = 153)	Remediated (n = 1,027)
ACC_PURCHASES= purchase to payment cycle.	23.8%	23.1%	23.4%	21.6%	25.8%
ACC_REVENUES = revenue to cash cycle.	19.7	17.8	20.8	48.4	19.0
ACC_COMPENSATION = compensation, payroll and benefits.	8.9	9.0	10.0	2.6	9.3

(continued on next page)

Panel C: Cycle/Account

	All ICD (n = 3,990)	Deficiencies (n = 2,328)	Significant Deficiencies (n = 461)	Material Weaknesses (n = 153)	Remediated (n = 1,027)
ACC_FIXEDINTANG = fixed assets and intangibles.	5.5	6.5	2.8	2.0	4.9
ACC_TAXES = taxes.	3.1	2.8	6.3	3.9	2.1
ACC_LIABILITIES = commitments and contingencies.	2.2	2.0	1.7	3.9	2.6
ACC_TREASURY = treasury and investments.	2.5	2.0	2.2	3.3	3.8
Total account-specific controls	64.2%	62.2%	66.6%	80.9%	65.3%

This table presents characteristics of individual ICD, overall and by severity classification. Percentages are computed on numbers of ICD for which we have data on the relevant variable. In some columns, percentages add to greater than 100 due to multiple categorizations of individual ICD.

highly skewed distribution. The mean number of SD for companies with ineffective (effective) controls is 18.27 (3.07), a significant difference ($p = 0.030$). The overall mean number of Deficiencies is 30.63.

Table 1, Panel B shows that auditors detected 72.1 percent of ICD. This statistic addresses our first research question, showing that auditors detect a large proportion of ICD even though clients are both documenting and testing controls. This percentage differs among severity classes of unremediated ICD ($p = 0.001$); it is highest for MW (84.3 percent) and lowest for deficiencies (74.7 percent). Panel B also shows that the percentage of misstatement-associated ICD increases with severity ($p < 0.001$), as does the percentage detected with substantive tests ($p < 0.001$). The percentage of ICD detected by control tests decreases with severity ($p = 0.042$).

Table 4, Panel A provides further evidence on detection method by documenting the types of ICD detected by auditors with substantive (control) tests. Column 1 shows that 81.5 percent of unremediated entity-level ICD are initially detected through control tests, while only 15.1 percent are detected through substantive tests. Within entity-level categories, control test detection rates are above 80 percent for all except general ledger controls, which include adjusting and closing entries. Account-specific ICD are also more likely to be detected though control tests, with tax accounts having the lowest of the listed percentages (64.8 percent overall). The only ICD type for which substantive and control test detection is roughly equivalent is misstatement-associated (45.0 versus 44.1 percent). Panel B shows that the percentage of misstatement-associated ICD detected with control tests is lower than for other ICD ($p < 0.001$). Columns 2 and 3 show a pattern of decreasing control test detection and increasing substantive test detection with increasing severity. Panel B shows that overall, the percentage of ICD detected with control tests decreases with severity ($p < 0.001$). Overall, this analysis shows that while more severe ICD are less likely to be detected through control tests, control tests still provide initial evidence for most MW. These results based on control-level data cast doubt on the effectiveness of relying on misstatements discovered by auditor substantive procedures to detect key flaws in controls over financial reporting, as suggested by KS from company-level data.

Client ICD Detection and Under-Classification

Table 5 presents results from estimating Model 1, examining factors associated with client ICD detection, for all unremediated ICD and for SD/MW only, along with descriptive statistics on model variables. Results for all unremediated ICD show that clients are less likely to detect ICD ultimately classified as MW or SD (both at $p = 0.011$) as well as entity-level ICD ($p = 0.024$). Client detection is more likely for larger clients ($p = 0.005$) and those using a large audit firm consultant ($p = 0.001$). However, independent reporting chain, internal audit function, IT integration, and longer client testing process are not associated with better client detection. Among SD/MW only, a large audit firm consultant again improves detection ($p = 0.005$) and entity-level ICD are less likely to be detected by clients ($p = 0.038$). In sum, Model 1 shows that clients are less likely to detect more severe and pervasive ICD, but use of large audit firm consultants and greater resources improves client detection.

According to AS No. 2, auditors should be aware of ICD detected by clients before beginning work in a given area of the audit. Because the auditor may also know the client's preliminary severity classification, we compare client and auditor ICD classification in these situations. Table 6, Panel A shows that 71.4 (64.8) percent of client-detected MW (SD) were initially under-classified by clients. These percentages are both significantly different from chance ($p = 0.005$ and $p = 0.023$, respectively), implying that auditors tend to override client classifications in a significant number of cases. Client under-classification could be due to lack of knowledge of the impact of specific controls, or an attempt to influence the auditor's classification to a less severe level.

TABLE 4

Supplemental Analysis of Detection Method for Auditor-Detected ICD

Panel A: Percent Detected with Control (Substantive) Tests, by ICD Characteristics

	1. All Unremediated	2. SD and MW	3. MW
Overall	82.1	72.7	66.7
	(14.3)	(22.8)	(28.7)
Entity Level	81.5	74.1	63.3
	(15.1)	(22.3)	(33.3)
ENTITY_CONTROLENVIR	83.7	88.6	100.0
	(9.9)	(2.9)	(0.0)
ENTITY_MONITORING	83.3	77.3	60.0
	(14.5)	(21.2)	(40.0)
ENTITY_GENERAL_LEDGER	69.7	55.8	56.8
	(26.2)	(39.5)	(37.8)
ENTITY_ITGC	93.3	96.6	80.0
	(5.0)	(2.6)	(20.0)
ENTITY_RISKASSESS	100.0	100.0	100.0
	(0.0)	(0.0)	(0.0)
Account-Specific	79.1	66.1	64.4
	(17.3)	(30.1)	(32.7)
ACC_PURCHASES	81.6	71.0	72.0
	(14.9)	(23.4)	(20.0)
ACC_REVENUES	79.1	72.0	66.2
	(16.5)	(25.0)	(30.8)
ACC_COMPENSATION	84.0	64.9	66.7
	(13.3)	(35.1)	(33.3)
ACC_TAXES	64.8	46.9	40.0
	(29.7)	(40.6)	(60.0)
Misstatement	45.0	30.6	35.9
	(44.1)	(57.3)	(48.7)

Panel B: Tests of Significance of Difference in Detection Method by ICD Characteristics

	Percent Detected with Control Tests	$\chi^2_{df=1}$
MW and SD	72.7%	39.5***
Deficiencies	84.9%	
Misstatement Associated	45.0%	341.1***
All other ICD	88.2%	

*** Indicates significance at ≤0.01.

This table presents supplemental analysis of detection method, among those ICD detected with auditor involvement, for which that information is available.

Panel A shows percentages of control (substantive) test discovery for all unremediated ICD (Column 1), SD/MW (Column 2), and MW (Column 3). Percentages may not add to 100 due to other possible methods of detection.

Panel B provides tests of detection method by severity, and between misstatement and no-misstatement subsamples.

TABLE 5

Descriptive Statistics and Model Results on Client Detection

	Percent of Unremediated ICD Detected by Clients Alone	Results of Model 1	All Unremediated ICD	SD/MW
Overall	23.9%	*CONSULTANT_BIG6* (+)	1.306***	1.735***
No Section 404 consultant	17.2	*CONSULTANT_OTHER* (+)	0.240	0.613
Smaller audit firm or other consultant	22.5	*REPORT_INDEPENDENT* (+)	0.171	0.358
Big 6 firm consultant	37.2	*INTERNAL_AUDIT* (+)	−0.588†	0.157
Section 404 results reported through management	23.4	*IT_INTEGRATION*	−0.330	−0.335
Section 404 reporting chain independent of management	24.6	*MONTHS_PRIOR_CLIENT* (+)	−0.038	0.078
No internal audit function	26.3	*MW* (−)	−0.827**	−0.195
Internal audit function	20.7	*SD* (−)	−0.472**	
Deficiencies	25.3	*ENTITY* (−)	−0.348**	−0.707**
SD	19.1	*LNASSETS* (+)	0.408**	0.319
MW	15.7	*FYE_2004* (−)	−0.517*	−0.329
Entity level	20.2	Constant	−7.574**	−6.272**
Not entity level	28.5	n	2,930	612
		Pseudo-R^2	0.065	0.082

*, **, *** Indicate significant effects at ≤0.10, ≤0.05, and ≤0.01, respectively, probability levels are one-tailed for directional expectations indicated in the table. Values in the cells are model coefficients.

† Indicates significance in the opposite direction to that predicted, at p ≤0.10 (two-tailed).

This table presents descriptive statistics and results of estimating Model 1, a logistic regression model (estimated with clustering on engagement ID) of factors associated with client detection of ICD without auditor assistance. Independent variables are defined in Tables 1 and 2.

TABLE 6

Client Under-Classification of ICD Severity

Panel A: Overall Statistics on Client Under-Classification

	SD	MW
ICD detected by clients, for which the initial client classification is known	91	28
Number (percent) for which the initial client classification was "deficiency"	59 (64.8%)	16 (57.1%)
Number (percent) for which the initial client classification was "SD"		4 (14.3%)
Overall number (percent) under-classified	59 (64.8%)	20 (71.4%)
$\chi^2_{df=1}$	8.011***	5.143**

Panel B: Descriptive Statistics and Model Results on Factors Affecting Client Under-Classification of SD and MW

% of SD/MW Under-Classified by Clients

High/significant control reliance	6.1
Other control reliance	89.5
No Section 404 consultant	20.0
Smaller audit firm or other consultant	70.8
Big 6 firm consultant	70.3
Reporting chain through management	78.9
Reporting chain independent of management	16.7
No internal audit function	76.5
Internal audit function	52.9
SD	64.8
MW	71.4
Entity level	48.1
Not entity level	80.6

Logistic Regression Model of Under-Classification

CONSULTANT_BIG6 (−)	0.968
CONSULTANT_OTHER (−)	1.214
REPORT_INDEPENDENT (−)	−18.603***
INTERNAL_AUDIT (−)	−0.107
IT_INTEGRATION (−)	0.186
MW (+)	17.875***
ENTITY	−1.059**
LNASSETS	−1.388
FYE_2004	0.468
Constant	27.990
n	119
Pseudo-R^2	0.513

(continued on next page)

, * Indicate significant effects at ≤0.05 and ≤0.01, respectively, probability levels are one-tailed for directional expectations indicated in the table. Values in the cells are model coefficients.

This table presents supplemental analysis of client under-classification for those SD/MW for which a preliminary client classification is available. The model is a logistic regression estimated with clustering on engagement ID, whose dependent variable = 1 if the client's classification is lower than the auditor's final classification, and 0 if the same. We include descriptive statistics on *IC_RELIANCE*, but this variable is omitted from the model as its presence prevents estimation.

Table 6, Panel B presents supplemental statistics on client classification of ICD whose final classification is SD or MW, and results of estimating a logistic regression model whose dependent variable (*UNDER_CLASSIFICATION)* equals 1 if the client's preliminary classification is lower than the auditor's final classification, and 0 otherwise. We expect that clients might not recognize, or might seek to downplay, more severe (MW) or pervasive (entity-level) problems. Consistent with this expectation, the coefficient on *MW* is positive and significant ($p < 0.001$). However, entity-level ICD are unexpectedly less likely to be under-classified ($p = 0.035$, two-tailed).[22] We also expect that clients with stronger Section 404(a) processes should be less likely to under-classify severity. Model results show that under-classification is less likely when the client has an independent reporting chain ($p < 0.001$), consistent with this expectation. Although Model 1 shows that large firm consultants improve ICD detection, Table 6 shows that independence in the reporting process plays a greater role in accurate client severity classification.[23]

Results of Estimating Models 2 and 3: Auditors' Severity Classification Decisions

Table 7 presents results on factors affecting auditors' severity classification decisions, investigated using Model 2 (distinguishing SD/MW from Deficiencies) and Model 3 (distinguishing MW from SD). These models have a Pseudo-R^2 of 0.352 and 0.437, respectively.

Client Characteristics and Section 404 Processes

Model 2 results show a negative and significant coefficient on *IC_RELIANCE* ($p = 0.013$) and a positive and significant coefficient on *NUM_MW_SQRT* ($p < 0.001$). Both imply that, when the client's overall control quality is lower, an ICD is more likely to be classified as SD relative to a deficiency. Concerning Section 404 process strength, Model 2 shows a positive coefficient on use of a large accounting firm as consultant ($p < 0.001$), but not for other consultants.[24] SD/MW classification is also more likely when Section 404 reporting is independent of management ($p = 0.001$). The coefficient on *IT_INTEGRATION* in Model 2 is negative and significant ($p = 0.012$), implying that better underlying IT controls are associated with less severe ICD. The presence of an internal audit function is not significant. Model 3 results show factors that contribute to distinguishing MW from SD. Regarding the overall controls context, *NUM_MW_SQRT* is again significant ($p < 0.001$), but *IC_RELIANCE* is not. An independent reporting chain is also positive and significant in Model 3 ($p = 0.037$).

Auditor Section 404 Processes

Model 2 results show that as expected, ICD detected by auditors are more likely to be SD/MW ($p = 0.008$). The length of the auditor testing period is also positive and significant ($p = 0.003$), suggesting that auditors who take more time to consider consequences of ICD classify them as more severe. ICD detected with a current misstatement are more likely to be SD/MW ($p < 0.001$), as are those discovered through substantive tests ($p = 0.074$). In Model 3,

[22] While Model 1 shows that clients are less likely to detect entity-level problems, this result implies they are more aware of their significance when detected.

[23] Although the descriptive statistics in Panel B include *IC_RELIANCE*, this variable is not included in the model, as its presence prevents estimation. This is likely due to the very low extent of under-classification among well-controlled companies. Table 5, Panel B shows that only 6.1 percent of ICD of clients of with higher control reliance are under-classified, relative to 89.5 percent of other clients.

[24] In Model 1 we find that large accounting firm consultants are associated with better ICD detection, and in Model 2 we find they are associated with higher severity classifications. These results could be explained by a tendency for poorly controlled companies to hire Big 6 consultants, but we find no association of use of large firm consultants with internal control reliance ($p = 0.555$). Model 2 also shows that an independent reporting chain is associated with higher severity classifications, which could be due to poorly controlled companies designing their processes in that manner. However, we also find no association of independent reporting with internal control reliance ($p = 0.869$). Thus, the more likely explanation is that use of those processes leads to higher quality Section 404(a) outcomes.

TABLE 7

Models 2 and 3: Control, Process, and Company Factors Associated with Severity Classifications

	Model 2: SD and MW versus Deficiencies	Model 3: MW versus SD
Client Characteristics and Detection Processes		
IC_RELIANCE (−)	−0.945**	0.086
TOTAL_MW_SQRT (+)	0.308***	0.763***
CONSULTANT_BIG6	1.892***	−1.473
CONSULTANT_OTHER	0.244	0.257
REPORT_INDEPENDENT (+)	1.464***	3.224**
INTERNAL_AUDIT	−0.217	2.266
IT_INTEGRATION	−0.527**	0.705
Auditor Detection Processes		
DETECTED_AUDITOR (+)	0.449***	0.240
MONTHS_PRIOR_AUDITOR (+)	0.324***	−0.680
MISSTATEMENT (+)	0.849***	0.636
DETECTED_SUBS (+)	0.559*	0.991*
ICD Types		
ENTITY_CONTROLENVIR (+)	0.434	2.603***
ENTITY_MONITORING (+)	0.228	−0.744††
ENTITY_INFOCOM (+)	−1.612††	
ENTITY_RISKASSESS (+)	−1.739†††	
ENTITY_GENERAL_LEDGER (+)	0.425**	0.433
ENTITY_ITGC	0.445*	−1.113*
ACC_REVENUES (+)	0.933***	2.060***
ACC_TAXES (+)	0.851**	0.839
ACC_PURCHASES	0.392	0.811
ACC_COMPENSATION	0.191	
ACC_FIXEDINTANG	−0.755*	
ACC_LIABILITIES	−0.208	
ACC_TREASURY	0.496	
DESIGN (+)	1.183***	0.993***
INEFFECTIVE (+)	0.760***	0.928***
INSUFFICIENT_TESTING (+)	0.627***	0.817**
Control Variables: Client Characteristics		
LNASSETS (−)	−0.245	0.584
LEVERAGE (+)	−0.563	1.980
CFOTA (−)	3.715††	−6.717**
ACCRUALS (+)	11.817***	−0.074
ARINVTA (+)	−4.915†††	7.002**
SEGMENTS (+)	−0.249†††	0.031
LOCATIONS (+)	−0.009	−0.232
NEW_PUBLIC (+)	0.848***	3.777**
FYE_2004 (−)	−0.361	−0.093
INDUSTRY_MFG	−0.979***	−1.488
Constant	3.655	−22.423
n	2,909	611
Pseudo-R^2	0.352	0.437

(continued on next page)

TABLE 7 (continued)

*, **, *** Indicate significant effects at ≤0.10,≤0.05, and ≤0.01, respectively, probability levels are one-tailed for directional expectations indicated in the table. Values in the cells are model coefficients.
††, ††† Indicate significance in the opposite direction to that predicted, at p ≤0.05 and p ≤0.01, respectively (two-tailed).
This table shows results of estimating Models 2 and 3, logistic regressions estimated using clustering on engagement ID. Model 2 explains factors differentiating deficiencies from SD (i.e., the ICD meets criteria for disclosure at least to the audit committee), and Model 3 explains factors differentiating SD from MW (i.e., the ICD meets criteria for public disclosure). Independent variables are defined in Tables 1 and 2. A blank cell in Model 3 results indicates that a variable is not included because its presence prevents estimation, due to low representation.

we again find that substantive test detection is associated with MW classification ($p = 0.062$). However, an associated misstatement does distinguish MW from SD, when controlling for the nature of the ICD and other variables in the models.

ICD Types

Our models also examine severity classifications of specific types of entity-level ICD. Model 2 results show differing associations across types. Coefficients on general ledger controls ($p = 0.036$) and ITGC ($p = 0.092$) are positive and significant. However, COSO categories of Information/Communications and Risk Assessment are negative and significant ($p = 0.034$ and $p < 0.001$, respectively, two-tailed), and Control Environment is not significant. In contrast, Model 3 shows that location in the Control Environment is a strong factor in distinguishing MW from SD ($p = 0.001$). Monitoring ICD are less likely to be MW than SD ($p = 0.054$), as are ITGC ($p = 0.085$).

We also examine the severity of specific account-level ICD types. Both revenue and tax accounts have positive coefficients in Model 2 ($p = 0.002$ and $p = 0.026$, respectively), consistent with regulators' attention to those areas prior to and during the study period. The revenue category is also significant in Model 3 ($p < 0.001$), likely due to the high volume of dollars flowing through that account. Model 2 results on the nature of detected ICD show positive and significant coefficients for design flaws ($p < 0.001$), ineffective controls ($p < 0.001$), and insufficient testing ($p = 0.003$). In Model 3, coefficients are again positive and significant for design flaws ($p = 0.003$), ineffective controls ($p < 0.001$), and insufficient testing ($p = 0.047$).

Control Variables

Models 2 and 3 also contain control variables for client characteristics. Model 2 results show that SD/MW are more likely for companies with lower earnings quality as measured by higher accruals ($p < 0.001$), and for new public companies ($p = 0.005$). SD/MW are unexpectedly less likely in more complex companies as measured by segments ($p < 0.001$) and concentration of assets in receivables and inventory ($p < 0.001$). Model 3 shows that MW classification is more likely in firms with lower cash flows ($p = 0.035$), greater complexity as measured by *ARINVTA* ($p = 0.047$), and in new public companies ($p = 0.030$).

Severity Classification with versus without an Associated Misstatement

Because prior research suggests differences in auditor judgment processes based on objectivity of evidence, we investigate the role of existing misstatements in severity classification by

estimating Models 2 and 3 in subsamples of ICD with/without an associated misstatement.[25] Table 8 compares coefficients on Section 404 process and ICD type variables in Models 2 and 3, across subsamples. We focus our discussion on variables with a significant cross-model comparison, and those that are significant in both models. In Model 2, all significant differences between subsamples are in the expected direction (i.e., a stronger association in the no-misstatement subsample). *NUM_MW_SQRT* is significant in both subsamples ($p = 0.059$ and $p < 0.001$, respectively), but the coefficient is statistically stronger in the no-misstatement subsample ($p = 0.029$). Thus, the client's ICFR context plays a larger role in severity classification when a misstatement has not occurred. Section 404 process variables are significant only in the no-misstatement subsample: *REPORT_INDEPENDENT* ($p = 0.002$), with the difference across models significant at $p = 0.010$, *IT_INTEGRATION* ($p = 0.003$), and *MONTHS_PRIOR_AUDITOR* ($p = 0.004$). Differences in coefficients for these three variables are significant across models with probabilities of at least $p < 0.05$. No ICD type variable is significantly different across subsamples, but both models show positive coefficients for revenues (misstatement: $p = 0.097$; no misstatement: $p < 0.001$) and design flaws (misstatement: $p = 0.022$; no misstatement: $p < 0.001$), indicating the pervasive importance of these ICD types.

Table 8 also shows results of comparing coefficients of Model 3 between misstatement and no-misstatement subsamples. *NUM_MW_SQRT* is again associated with MW classification in the no-misstatement subsample ($p = 0.004$) and its association is statistically greater than when there is a misstatement ($p = 0.010$). The same pattern holds for *REPORT_INDEPENDENT* ($p = 0.017$) and *DETECTED_SUBS* ($p = 0.002$). However, *IC_RELIANCE* is significant only when there has been a misstatement ($p < 0.001$), and the association is stronger in that subsample ($p = 0.010$). The latter result implies that when a misstatement amount is available, the auditor is less likely to judge the implications of the ICD to be material in clients for which the audit engagement was performed with the perception that overall controls were good. However, control reliance is not associated with MW classification when there is not a misstatement.[26] Among ICD type variables, revenue ICD are more likely to be MW in both subsamples ($p < 0.001$ and $p = 0.015$, respectively). General ledger ICD are significant only in the no-misstatement subsample ($p = 0.001$), and the test of association between subsamples is significant ($p = 0.039$).

Taken together, Table 8 results show that client and auditor process measures and characteristics are more likely to influence severity classification when the control has not yet failed to prevent/detect a misstatement. This pattern of findings is consistent with prior research on the impact of imprecise standards (Mayhew et al. 2001; Nelson et al. 2002; DeZoort et al. 2003), and implies that when a control failure is not yet documented, auditors use a wider variety of other information about the control flaw and its context to make and justify severity decisions.

V. CONCLUSIONS AND LIMITATIONS

This study contributes to the literature on ICFR by examining the relative effectiveness of companies and their auditors in Section 404 ICD detection, the processes used by those parties,

---

[25] Due to the difficulty of interpreting interactions in logit, we compare coefficients between subsamples using seemingly unrelated estimation in Stata. Our cross-model comparisons are limited in that the presence of some variables prevents estimation in certain subsamples (e.g., consultant use in Model 3).

[26] There are several possible reasons for this pattern, which our data cannot distinguish. One is that clients with high control reliance might have more effective compensating or monitoring controls, which would mitigate a MW classification. This could be due to auditors anchoring on the available misstatement amount (not on the potential amount of future misstatements) and judging that amount to be less severe for a client with good controls. A further reason to downplay the severity of an ICD with a detected misstatement for an engagement in which controls have been relied upon, is that if a MW is found, the whole audit strategy of reducing substantive testing due to good controls is called into question.

TABLE 8

Comparison of Model Coefficients in Subsamples With/Without an Associated Misstatement

	Model 2: (SD/MW versus Deficiencies)			Model 3: (MW versus SD)		
	Misstatement Subsample	No Misstatement Subsample	Cross-Model Comparison ($\chi^2_{df=1}$)	Misstatement Subsample	No Misstatement Subsample	Cross-Model Comparison ($\chi^2_{df=1}$)
IC_RELIANCE (−)	−0.727**	−0.808	0.00	−5.580***	2.264	9.80***
TOTAL_MW_SQRT (+)	0.134*	0.351***	3.61**	0.060	1.517***	5.43***
CONSULTANT_BIG6	0.879	1.913***	1.26	−5.178		
CONSULTANT_OTHER	−0.616	0.295	3.06*	−3.370		
REPORT_INDEPENDENT (+)	0.027	1.950***	5.48**	−1.409	6.003**	2.21*
INTERNAL_AUDIT	−0.687*	−0.113	0.64			
IT_INTEGRATION	0.287	−0.774***	7.86***	−0.285	2.091	0.35
DETECTED_AUDITOR (+)	−0.098	0.499***	0.81	−0.162	0.071	0.17
MONTHS_PRIOR_AUDITOR (+)	−0.002	0.371***	3.75**	0.472	−0.183	0.23
DETECTED_SUBS (+)	0.652**	0.391	0.19	0.049	3.762***	8.66***
ENTITY_CONTROLENVIR (+)	−0.183	0.542	0.42		3.661***	
ENTITY_MONITORING (+)	0.431	0.242*	0.13	0.183	−1.496†††	1.94
ENTITY_GL (+)	0.236	0.429*	0.12	−0.328	1.512***	3.10**
ENTITY_ITGC (+)		0.506**				
ACC_REVENUES (+)	0.615*	1.140***	0.68	2.974**	2.822***	0.02
ACC_TAXES (+)	1.461**	0.360	1.85			
DESIGN (+)	1.047**	1.102***	0.02	0.874	0.752**	0.00
INEFFECTIVE (+)		0.594**			0.709***	
INSUFFICIENT_TESTING (+)		0.621***			1.263***	

*, **, *** Indicate significant effects at ≤0.10, ≤0.05, and ≤0.01, respectively, probability levels are one-tailed for directional expectations indicated in the table. Values in the cells are model coefficients.
This table shows partial results of estimating Models 2 and 3 in sub-samples of ICD with/without an associated misstatement.
(The complete models are estimated but for efficiency, only the test variables are tabled.) Independent variables are defined in Tables 1 and 2.

and factors influencing ICD severity classification. Our data provide the only systematic control-level evidence available on the baseline set of detected ICD. Of those, about one quarter were remediated prior to year-end, showing progress in improving controls not visible to the public. Of all ICD detected, fewer than 4 percent are publicly disclosed as MW, but many companies having a "clean" Section 404 report have at least one SD—a control flaw serious enough to report to the audit committee. These results demonstrate the importance of looking beyond publicly available information to capture the full picture ICFR problems detected under Section 404.

Prior research does not differentiate ICD detection by auditors and clients (Leone 2007). Our data show that auditors detect almost three-fourths of ICD, especially those that are more severe and pervasive. Importantly, this low level of client detection occurs when clients are aware that auditors will soon follow with their own tests. Our results also reveal company characteristics associated with better ICD detection, including use of large audit firm consultants, Section 404(a) reporting independent of management, and greater resources (company size). While we do not find the mere presence of an internal audit function improves detection, Lin et al. (2011) find that characteristics of internal auditors do influence the quality of companies' Section 404 activities.

Our results also shed light on auditor detection processes, a controversial issue at the heart of the debate on the value of Section 404(b). We find that two-thirds of auditor-detected MW are signaled through control tests. Therefore, without auditor control testing, many key flaws in ICFR could have been missed (e.g., over 80 percent of ICD in Control Environment and revenues). KS present a contrasting perspective, based on Audit Analytics data indicating that Section 404 opinions of companies with ineffective controls often mention misstatements. KS therefore suggest that substantive procedures, enhanced by "tracing back" from misstatements to the underlying control failure, could be an effective surrogate for Section 404 testing in identifying most companies with ineffective controls.[27] There are several reasons underlying the difference in conclusions between their research and ours. Audit Analytics data are organized at the company level, not at the level of the individual MW. Thus, KS's results reveal that misstatements are associated with at least one of a company's MW, but not which or how many MW caused a misstatement.[28] Our data augment publicly available information by revealing the number of individual control flaws detected per company, and how each was detected. Thus, we show the degree to which a policy not relying on auditor testing of controls would miss important control flaws. Our data show that, for 27 percent of our companies with ineffective ICFR, no MW are associated with a misstatement, and in an additional 40 percent, one or more MW are not associated with a misstatement. This implies that many MW would likely be missed if relying solely on detected misstatements to identify control deficiencies. Our data additionally show that most SD would not be identified through detected misstatements. In sum, we conclude that relying on misstatements to identify ICD will result in missing many severe control flaws. Thus, the intended overall purpose of SOX Section 404 (i.e., to improve financial reporting) would not be completely fulfilled for those companies.

In addition to the difference in unit of analysis (company or ICD), our study differs from KS in the purpose of Section 404. They focus on disclosure of ineffective controls in the auditor's opinion as that Section's purpose. This view implies that if disclosure of ineffective controls (i.e.,

[27] Our data show a similar proportion of companies with ineffective controls for which at least one MW is associated with a misstatement. KS cite a range of 57 to 73 percent based on the aggregated data; in our sample this proportion is 73 percent.

[28] Additionally, reliance on "tracing back" from misstatements to identify related control flaws assumes that the yield of misstatements detected from substantive tests would be the same without the information provided by the preceding extensive control tests required under Section 404. This seems unlikely, given that control tests should direct the auditor's attention to areas of greater risk.

at least one MW) was obtained for a company through an alternative approach, then a purpose of Section 404 would be fulfilled for that company. However, the purpose of Section 404 requirements is to identify and correct control deficiencies and improve the reliability of financial reporting (e.g., COSO 2006; PCAOB 2004). Thus, an ineffective controls opinion is a consequence of the process of controls assessment, but not its main purpose. In sum, publicly available data are limited in supporting a conclusion on the efficiency and effectiveness of the controls assessment and testing process in improving financial reporting reliability. For the reasons cited above, our data provide a more complete picture of the results of Section 404 activity.

We also study how auditors classify ICD in the context of imprecise standards. For auditors, this situation is analogous to imprecise financial reporting standards, which have been shown to decrease auditors' ability to prevail in booking proposed audit adjustments (Nelson et al. 2002). Our results show that ICD associated with an existing misstatement have higher severity classifications. It is likely that a current control failure makes future failures easier to visualize, and provides more objective evidence to justify severity to clients (e.g., Kennedy et al. 1997). Nevertheless, our data show that a misstatement is neither necessary nor sufficient for high severity classification, prompting our investigation of how auditors make severity judgments when the control has *not yet failed*. For these ICD, results generally imply a stronger association of task expertise and process independence, relative to those ICD that have failed in the current period. Use of a large accounting firm consultant, lower quality of internal control, better client IT integration, and auditor substantive test detection are more strongly associated with severity when the objective evidence provided by a current misstatement is unavailable.

Our investigation of the nature of ICD also contributes to the literature by showing which types of entity-level and account-level controls that auditors consider could produce future misstatements most threatening to financial reporting quality. Within entity types, it is interesting to compare our findings on COSO Control Environment and Risk Assessment categories. Relatively few Control Environment ICD are detected, but those are more likely to be classified as MW. In contrast, COSO Risk Assessment ICD are infrequently found, and are *less* likely to be classified as severe. While this might imply that companies' risk assessment controls were of high quality during the study period, evidence from the recent financial crisis suggests that many companies did not effectively monitor their risks. Further research should examine whether companies and auditors have increased their focus on the quality of risk assessment controls in the current environment.

Among specific accounts/cycles, revenue ICD are more likely to be severe, consistent with the importance of revenues in equity valuation. We also find that tax ICD are more likely to be classified as at least SD. This is consistent with auditors' concerns that many clients during the study period were unable to perform the tax accrual without assistance. Regarding the nature of control failures, we find that ineffective controls (well designed but not functioning) are most frequent. Those ineffective controls, as well as design flaws (comprising missing controls as well as controls that do not meet their intended purpose), and controls that have been insufficiently tested by clients, are all considered more severe relative to documentation problems. Further research should investigate whether the incidence of these more severe types of control flaws has declined following several years of ICD remediation and retesting under Section 404.

Our study complements a behavioral experiment by Earley et al. (2008), which finds that clients' classifications influence auditors' judgments. Our archival data show that clients' classifications are overwhelmingly lower than those of auditors, implying that auditors often override

clients' views.[29] Client under-classification could be caused by differing incentives, uncertainty regarding appropriate procedures (due to lack of precise SEC guidance for company managements during our study period), and/or lack of knowledge of internal controls. The first explanation seems consistent with our finding that clients with a more independent Section 404(a) process are less likely to under-classify, but future research should investigate these and other possible causes.

The generalizability of this study's findings may be limited by several features of its design. First, our conclusions are limited by restricting the sample to smaller accelerated filers that more closely apply to non-accelerated filers and foreign companies not currently subject to Section 404 or similar regulation. Second, engagements in our sample were performed under AS No. 2. Additional guidance on company and auditor processes for Section 404 were subsequently issued (SEC 2005; 2006), and AS No. 5 (PCAOB 2007) superseded AS No. 2. Both the revised auditing standard and related SEC guidance were directed toward helping companies and auditors achieve greater efficiencies but not to "lower the bar" for reporting issues. While the current study's results should have continuing relevance, research in the AS No. 5 environment would provide insight on effects of application of that standard. Third, it is difficult to directly compare the proportions of MW identified in our sample with public disclosures, as individual control flaws may be aggregated in those reports. Fourth, sample engagements were chosen by participating firms. While our sample represents publicly available data at the company level (as indicated by statistical comparisons with public data reported above), we cannot rule out the possibility that nonrandom selection might have influenced our results.

In sum, this study provides a unique view into processes and outcomes of the first two years of Section 404 compliance. In terms of process, our results support the value of auditor involvement at two stages of the ICFR assessment process (detection and classification), and contribute to the understanding of factors associated with client and auditor performance in both stages. Regarding outcomes, we show that even companies with a clean Section 404 auditor's opinion can have numerous and/or significant ICD. Having been identified, those ICD may now be corrected, thus improving financial reporting quality. Overall, results of this study imply that the recently enacted exemption of non-accelerated filers from SOX Section 404(b), which requires auditor testing and an audit report on internal control, is likely to undermine the objectives of SOX for these entities, resulting in relatively less reliable financial reporting than had 404(b) provisions been implemented. Further, risk assessments in subsequent years may be less effective without the detailed information on client controls that comes to light when the independent auditor detects and classifies control flaws.

REFERENCES

Allen, R., D. Hermanson, T. Kozloski, and R. Ramsay. 2006. Auditor risk assessment: Insights from the academic literature. *Accounting Horizons* 20 (2): 157–177.

Amer, T., K. Hackenbrack, and M. Nelson. 1994. Between-auditor differences in the interpretation of probability phrases. *Auditing: A Journal of Practice & Theory* 13 (1): 126–136.

American Institute of Certified Public Accountants (AICPA). 1988. *The Auditor's Consideration of Internal Control Structure in the Financial Statement Audit.* SAS No. 55. New York, NY: AICPA.

———. 1997. *Consideration of Internal Control in a Financial Statement Audit: An Amendment to Statement on Auditing Standards No. 55.* SAS No. 78. New York, NY: AICPA.

[29] Prior research finds lower rates of MW disclosure under Section 302, which does not require companies to test controls (e.g., Hoitash et al. 2009; Hermanson and Ye 2009). The current study shows that clients under-classify severity even when they test controls.

Asare, S., and A. Wright. 2008. Equity analysts' reactions to type of control deficiency and likelihood threshold in adverse control reports. Working paper, University of Florida.

Ashbaugh-Skaife, H., D. Collins, and W. Kinney. 2007. Discovery and reporting of internal control deficiencies prior to SOX-mandated audits. *Journal of Accounting and Economics* 44 (1–2): 166–192.

———, D. Collins, W. Kinney, and R. LaFond. 2008. The effect of SOX internal control deficiencies and their remediation on accrual quality. *The Accounting Review* 83 (1): 217–250.

BDO Seidman LLP, Crowe Chizek & Company LLC, Deloitte & Touche LLP, Ernst & Young LLP, Grant Thornton LLP, Harbinger PLC, KPMG LLP, McGladrey & Pullen LLP, PricewaterhouseCoopers LLP, and W. Messier. 2004. *A Framework for Evaluating Deficiencies and Exceptions.* Version 3, December 20. New York, NY: BDO Seidman LLP et al.

Bedard, J., R. Hoitash, and U. Hoitash. 2007. Audit pricing and internal control disclosures among non-accelerated filers. *Research in Accounting Regulation* 20: 103–126.

Cohen, J., and D. Hanno. 2000. Auditors' consideration of corporate governance and management control philosophy in preplanning and planning judgments. *Auditing: A Journal of Practice & Theory* 19 (2): 133–146.

———, G. Krishnamoorthy, and A. Wright. 2002. Corporate governance and the audit process. *Contemporary Accounting Research* 19 (4): 573–594.

Committee of Sponsoring Organizations (COSO). 2006. *Internal Control over Financial Reporting—Guidance for Smaller Public Companies*. New York, NY: AICPA.

Curtis, M., D. Deis, G. Jenkins, and J. Bedard. 2009. Information systems and the financial statement auditor: A research synthesis. *Journal of Information Systems* 23 (1): 79–96.

DeZoort, T., D. Hermanson, and R. Houston. 2003. Audit committee support for auditors: The effects of materiality justification and accounting precision. *Journal of Accounting and Public Policy* 22 (2): 175–199.

Doyle, J., W. Ge, and S. McVay. 2007. Determinants of weaknesses in internal control over financial reporting. *Journal of Accounting and Economics* 44 (1–2): 193–223.

Earley, C., V. Hoffman, and J. Joe. 2008. Reducing management's influence on auditors' judgments: An experimental investigation of SOX 404 assessments. *The Accounting Review* 83 (6): 1461–1485.

Elder, R. A., and R. D. Allen. 2003. A longitudinal field investigation of auditor risk assessments and sample size decisions. *The Accounting Review* 78 (4): 983–1002.

Felix, W., A. Gramling, and M. Maletta. 2001. The contribution of internal audit as a determinant of external audit fees and factors influencing this contribution. *Journal of Accounting Research* 39 (3): 513–534.

Financial Accounting Standards Board (FASB). 1975. *Accounting for Contingencies*. Statement of Financial Accounting Standards No. 5. Stamford, CT: FASB.

Hammersley, J., L. Myers, and C. Shakespeare. 2008. Market reactions to internal control weakness disclosures and to the characteristics of those weaknesses under section 302 of the Sarbanes-Oxley Act of 2002. *Review of Accounting Studies* 13 (1): 141–165.

Hermanson, D., and Z. Ye. 2009. Why do some accelerated filers with SOX Section 404 material weaknesses provide early warning under Section 302? *Auditing: A Journal of Practice & Theory* 28 (2): 247–271.

Hoitash, R., U. Hoitash, and J. Bedard. 2008. Audit fees and internal control problems disclosed under the Sarbanes-Oxley Act. *Auditing: A Journal of Practice & Theory* 27 (1): 105–126.

Hoitash, U., R. Hoitash, and J. Bedard. 2009. Corporate governance and internal control quality: A comparison of regulatory regimes. *The Accounting Review* 84 (3): 839–867.

IT Governance Institute. 2004. *IT Control Objectives for Sarbanes-Oxley: The Importance of IT in the Design, Implementation and Sustainability of Internal Control over Disclosure and Financial Reporting*. Chicago, IL: IT Governance Institute.

Kennedy, J., D. Kleinmuntz, and M. Peecher. 1997. Determinants of the justifiability of performance in ill-structured audit tasks. *Journal of Accounting Research* 35 (Supplement): 105–123.

Kinney, W.R., and M. Shepardson. 2010. Do control effectiveness disclosures require internal control audits: A natural experiment with small U.S. public companies. Forthcoming. *Journal of Accounting Research.*

Krishnan, J., D. Rama, and Y. Zhang. 2008. Costs to comply with SOX Section 404. *Auditing: A Journal of Practice & Theory* 27 (1): 169–186.

Larcker, D. F., and T. O. Rusticus. 2010. On the use of instrumental variables in accounting research. *Journal of Accounting and Economics* 49 (3): 186–205.

Leone, A. 2007. Factors related to internal control disclosure: A discussion of Ashbaugh, Collins and Kinney (2007) and Doyle, Ge, and McVay (2007). *Journal of Accounting and Economics* 44 (1–2): 224–237.

Li, C., L. Sun, and M. Ettredge. 2010. Financial executive qualifications, financial executive turnover, and adverse SOX Section 404 opinions. *Journal of Accounting and Economics* 50 (1): 93–110.

Lin, S., M. Pizzini, M. Vargus, and I. Bardhan. 2011. The role of internal audit function in the disclosure of material weaknesses. *The Accounting Review* 86 (1).

Mayhew, B., W. J. Schatzberg, and G. Sevcik. 2001. The effect of accounting uncertainty and auditor reputation on auditor independence. *Auditing: A Journal of Practice & Theory* 20 (2): 49–70.

Messier, W., N. Martinov-Bennie, and A. Eilifsen. 2005. A review and integration of empirical research on materiality: Two decades later. *Auditing: A Journal of Practice & Theory* 24 (2): 153–187.

Nelson, M., J. Elliott, and R. Tarpley. 2002. Evidence from auditors about managements' and auditors' earnings management decisions. *The Accounting Review* 77 (Supplement): 175–202.

Ng, T., and H.-T. Tan. 2003. Effects of authoritative guidance availability and audit committee effort on auditor's judgments in an auditor-client negotiation context. *The Accounting Review* 78 (3): 801–818.

Public Company Accounting Oversight Board (PCAOB). 2004. *An Audit of Internal Control over Financial Reporting Performed in Conjunction with an Audit of Financial Statements*. Auditing Standard No. 2. Washington, D.C.: PCAOB.

———. 2007. *An Audit of Internal Control over Financial Reporting that is Integrated with an Audit of Financial Statements*. Auditing Standard No. 5. Washington, D.C.: PCAOB.

Securities and Exchange Commission (SEC). 1999a. *Materiality*. SEC Staff Accounting Bulletin No. 99. Washington, D.C.: SEC.

———. 1999b. *Revenue Recognition in Financial Statements*. SEC Staff Accounting Bulletin No. 101. Washington, D.C.: SEC.

———. 2005. Staff statement on management's report on internal control over financial reporting. (May 16). Washington, D.C.: SEC.

———. 2006. SEC announces next steps for Sarbanes-Oxley implementation. (May 17). Washington, D.C.: SEC.

U.S. Congress. 2002. *The Public Company Accounting Reform and Investor Protection Act of 2002 (The Sarbanes-Oxley Act)*. Public Law No. 107-204, 116 Statute 745 (July 30). Washington, D.C.: Government Printing Office.

———. 2010. *Dodd-Frank Wall Street Reform and Consumer Protection Act*. Public Law No. 111-203 (July 21). Washington, D.C.: Government Printing Office.

Wolfe, C., E. Mauldin, and M. Diaz. 2009. Concede or deny: Do management persuasion tactics affect auditor evaluation of internal control deviations? *The Accounting Review* 84 (6): 2013–2037.

THE ACCOUNTING REVIEW
Vol. 86, No. 3
2011
pp. 857–886

American Accounting Association
DOI: 10.2308/accr.00000037

Capital Structure, Cost of Capital, and Voluntary Disclosures

Jeremy Bertomeu
Northwestern University

Anne Beyer
Stanford University

Ronald A. Dye
Northwestern University

ABSTRACT: This paper develops a model of financing that jointly determines a firm's capital structure, its voluntary disclosure policy, and its cost of capital. Investors who receive securities in return for supplying capital sometimes incur losses when they trade their securities with an informed trader. The firm's disclosure policy and the structure of its securities determine the information advantage of the informed trader and, hence, the size of investors' trading losses and the firm's cost of capital.

We establish a hierarchy of optimal securities and disclosure policies that varies with the volatility of the firm's cash flows. Debt securities are often optimal, with the form of debt—risk-free, investment grade, or "junk"—varying with the firm's cash flow volatility. Though the model predicts a negative association between firms' cost of capital and the extent of information firms disclose, more expansive voluntary disclosure does not cause firms' cost of capital to decline. Mandatory disclosures alter firms' voluntary disclosures, their capital structure choices, and their cost of capital.

Keywords: *cost of capital; capital structure; voluntary disclosures.*

I. INTRODUCTION

We develop a model that jointly explains a firm's voluntary disclosure policy, its capital structure, and its cost of capital. While not previously recognized in the literature, it is intuitive that there should be a relationship between a firm's capital structure and its disclosure policy, because—barring agency problems—managers will choose their firm's disclo-

We thank seminar participants at the 2009 AAA Annual Meeting, Carnegie-Mellon University, Stanford University, University of Bern, University of California, Berkeley, University of California, Los Angeles, University of Pennsylvania, and Washington University in St. Louis, as well as Stan Baiman, Edwige Cheynel, Enrique de la Rosa, Pingyang Gao, Miles Gietzman, Jon Glover, Ilan Guttman, Jack Hughes, Pierre Liang, Ivan Marinovic, Paul Newman (editor), Tjomme Rusticus, Hyun Shin, Sri Sridhar, Phil Stocken, Amir Ziv, and two anonymous referees for comments on previous drafts of the manuscript. Professor Dye also thanks the Accounting Research Center at Northwestern University for financial support.

Editor's note: Accepted by Paul Newman.

Submitted: January 2009
Accepted: October 2010
Published Online: April 2011

sure policy so as to maximize the market's perceptions of the expected value of the firm-owners' residual claims. Hence, the form of those residual claims—equivalently, the form of the securities that the firm sells to investors—affects what managers maximize. In turn, the firm's disclosure policy and its capital structure choices affect the firm's cost of capital.

In the model we study, a firm receives capital by issuing securities that promise future cash payouts to investors. The central feature of the model is that the amount investors are willing to pay for the securities may be less than the expected present value of the securities' cash payouts because the investors anticipate at the time they buy the securities that they may bear trading losses if they have to liquidate the securities "early"—i.e., before maturity. These potential trading losses arise because the market in which the securities are traded is presumed to be populated by some superiorly informed traders. Compensating investors for these expected trading losses *ex ante* constitutes a cost the firm bears in raising capital: it is *the* source of the "cost of capital" in our model.

The firm's manager can reduce investors' expected trading losses by reducing the information asymmetry between the superiorly informed traders and the market maker who sets the price for the securities. This can be accomplished in two related ways. First, since the informed traders' private information about the firm is likely to overlap with the manager's private information, the manager can disclose his private information.[1] Second, the manager can have the firm issue securities whose value is not "informationally sensitive" (e.g., Sunder 2006), meaning that their value does not vary with the informed traders' private information. For example, were the securities risk-free debt, then the "informed" traders—regardless of their information—would have no information advantage over the market maker in assessing the securities' value. Thus, a firm's capital structure and its disclosure policy jointly determine how much information asymmetry remains among market participants, and this remaining information asymmetry determines investors' expected trading losses and, hence, the firm's cost of capital.

We develop a formula, applicable to all securities, that establishes how a firm's cost of capital depends on the security's upside potential (the difference between a security's payoff if the firm's realized cash flows assume "high" or "medium" values) and its downside risk (the difference between a security's payoff if the firm's realized cash flows assume "medium" or "low" values). The formula shows that the contribution of a security's upside potential and downside risk to the firm's cost of capital varies with the firm's disclosure policy. Specifically, the contribution of a security's upside potential to the firm's cost of capital decreases as the firm discloses more information, and, perhaps surprisingly, the contribution of a security's downside risk to the firm's cost of capital *increases* as the firm discloses more information.

Our model establishes that debt is always an optimal security and that there is a hierarchy among debt securities and optimal disclosure policies that varies with the volatility of the firm's cash flows. A firm with very low volatility in its cash flows prefers to raise capital by issuing risk-free debt and adopting an "expansive" disclosure policy.[2] As its cash flow volatility increases, the firm prefers to use investment-grade debt (that defaults with low probability) combined with the continued use of an expansive disclosure policy. Then, as its cash flow volatility increases still further, the firm will continue to use investment-grade debt, but it will curtail its disclosures and adopt a "limited" disclosure policy. As its cash flows become even more volatile, the firm will switch to using "junk" debt (that defaults if anything other than the highest cash flows occur),

[1] We conform to the AAA's style guidelines by mixing gender references in the following: specifically, when describing or referring to actors in the model, we arbitrarily assigned the gender of the manager and market maker to be male, and the gender of the investor and informed trader to be female.

[2] We defer the precise definitions of an "expansive" and a "limited" disclosure policy to the text below.

accompanied by limited disclosure. Finally, as its cash flow volatility becomes greater still, we demonstrate that it is impossible to finance the firm with any form of security accompanied by any form of disclosure.

While reminiscent of Myers and Majluf's (1984) famous financial hierarchy (inside financing is least expensive; outside financing with debt is more expensive; outside financing with equity is most expensive), our hierarchy is distinct from Myers and Majluf's (1984) hierarchy in two fundamental respects. First, our hierarchy combines a firm's capital structure choice with its voluntary disclosure policy, whereas Myers and Majluf (1984) make no reference to firms' disclosures. Second, our capital structure hierarchy is indexed to the volatility of a firm's cash flows, whereas Myers and Majluf's (1984) hierarchy is indexed to the relative costliness of alternative financing methods, holding cash flow volatility fixed.

We also explore how mandatory disclosure requirements interact with firms' voluntary disclosure decisions to affect firms' cost of capital and securities design choices. Some mandatory disclosure requirements are shown to inhibit firms from making voluntary supplementary disclosures, and other mandatory disclosure requirements are shown to encourage firms to make voluntary supplementary disclosures. In addition, we show that imposing mandatory disclosure requirements sometimes radically alters what securities firms wish to offer their investors. In particular, we show that it is sometimes optimal for the firm to raise capital by issuing equity rather than debt.

Related Literature

While links between a firm's disclosure policy and its cost of capital have been established in prior accounting research (e.g., Botosan 1997; Botosan and Plumlee 2002), and links between a firm's capital structure and its cost of capital separately have been identified through research in finance (e.g., in Myers and Majluf 1984), we believe this is the first study in accounting or finance that endogenously connects a firm's disclosure policy to its capital structure, and *a fortiori*, the first that connects all three of these components of a firm's financial structure.

Other models in accounting and finance have emphasized how information asymmetry affects firms' decisions to raise capital. For example, Korajczyk et al. (1991, 1992) noted that firms tend to raise new capital in periods where the information asymmetry between insiders of the firm and outside investors is small so as to eliminate the lemons' problem with issuing equity (noted originally by Myers and Majluf [1984]). Unlike Korajczyk et al. (1991, 1992), we are (1) concerned about not when firms raise capital, but rather in what form firms raise capital, and (2) not concerned about disclosures before the firm raises capital, but rather about disclosures after the firm has raised capital.[3]

Several recent studies discuss links between the quality of firms' financial statements and their cost of capital. For example, Jorgensen and Kirschenheiter (2003) show that firms have incentives to disclose information selectively that indicates lower cash flow risk, thus leading to low-risk premia conditional on voluntary disclosure. As another example, Christensen et al. (2002) develop a theory that explains when and how managers' compensation should be adjusted for the cost of capital.

Our model is also linked to the large literature in finance on securities design. For example, see Harris and Raviv (1991) for an early survey of this literature, and Nachman and Noe (1994), DeMarzo and Duffie (1999), Demange and Laroque (1995), Rahi (1995), Fulghieri and Lukin (2001), Sunder (2006), Bias et al. (2007), Casamatta (2003), and Repullo and Suarez (2004) for

[3] But, the present model and Korajczyk et al. (1991, 1992) share the important property that the disclosure of information has real effects, which is also true of a substantial part of the disclosure literature (e.g., Kanodia et al. 2000; Stocken 2000; Pae 2002).

more recent work. As far as we are aware, none of the prior research on securities design has emphasized the interdependencies among securities design, cost of capital, and voluntary disclosures, which is our focus.

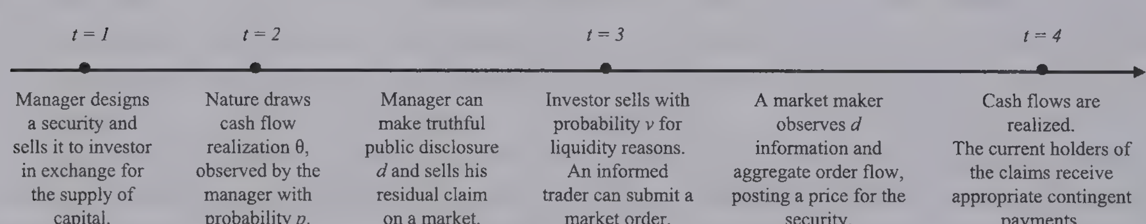
II. MODEL SETUP

The timeline for the model consists of four steps, presented in Figure 1. At step $t = 1$, an individual (called the "manager") owns a production technology (called the "firm") that converts capital invested in the first step stochastically into cash flows $\tilde{\theta}$ in the fourth step. The manager, who has no capital, offers (one unit of) a security $\tilde{\psi}$ to an investor in return for the investor supplying capital to the firm.

At step $t = 2$, with probability $p \in (0,1)$ the manager privately learns the realization of the firm's cash flows, and with probability $1 - p$, the manager learns nothing about those cash flows. If the manager receives no information, then he necessarily makes no disclosure. But, if he does receive information, then he may either disclose that information or disclose nothing, based on whether the disclosure or nondisclosure increases the capital market's perceptions of the expected value of the manager's residual claim $\tilde{\theta} - \psi(\tilde{\theta})$.[4] We assume all disclosures are truthful, but the manager cannot credibly disclose that he has received no information. We let $d(\theta) = \varnothing$ indicate that the manager makes no disclosure and $d(\theta) = \theta$ indicate that the manager discloses that the cash flows are θ.

At step $t = 3$, a market for the security $\tilde{\psi}$ opens.[5] With probability v, the investor experiences a liquidity shock that forces her to sell her entire interest in security $\tilde{\psi}$ on this market. We let $q = 1$ ($q = 0$) indicate that the investor is (is not) subject to a shock, so $\Pr(\tilde{q} = 1) = v$.[6] If not subject to a shock, though the investor may still sell her security, it will be clear in what follows that she will elect not to. The market clearing price of security $\tilde{\psi}$ is then established via a Kyle (1985)-type market maker, who sets the security's market price based on knowledge of the aggre-

FIGURE 1
Timeline



[4] We take the assumption that the manager seeks to maximize the expected value of his residual claim as exogenous. In an earlier working paper, we endogenize the objective function. The working paper version also extends the technology to two required inputs, capital and "effort" supplied by the manager.

[5] In a working paper version, we had two securities markets open, one for the security $\tilde{\psi}$ and one for the manager's residual security $\tilde{\theta} - \tilde{\psi}$. The results obtained by adding this second securities market are the same as in the present model.

[6] Our results do not depend on the assumption of an asymmetry in the frequency of the investor's and manager's liquidity shocks. What is critical is not how frequently the manager faces a liquidity shock *vis-à-vis* the investor, but rather that, whether subject to a liquidity shock, the manager decides whether to disclose his private information (when he receives it) based on whether the disclosure increases his firm's expected selling price.

gate order flow in the security. This aggregate order flow potentially includes orders from an informed trader. While the informed trader may go "long" in security $\tilde{\psi}$, we presume she is prohibited from "shorting" the security.[7]

Formally, we represent the informed trader as one who privately knows, when the manager makes no disclosure, whether the reason for the manager's nondisclosure was due to the manager not having received information or having deliberately withheld information he did receive.[8] In the latter case—where the investor knows that the manager withheld information—we presume that the trader does not know the value of the information the manager received. This constitutes the source of the informed trader's informational advantage over the market maker; when the manager makes no disclosure, the market maker, unlike the informed trader, does not know the reason for the manager's nondisclosure. This formal representation is intended to depict one of many possible ways an informed trader knows less than the manager. Results similar to those we obtain would hold for other representations where the informed trader's information is a coarsification of the manager's information.

At step $t = 4$, the firm's cash flows occur, and the holders of the security at this time are paid off according to the terms specified by the security.[9]

Having completed the description of the timeline for the model, we conclude this introduction by offering more specifics about both the firm's stochastic production technology and the security the manager offers investors. The production technology requires a fixed amount of capital, which we scale to be \$1, to be supplied by an investor. Less than \$1 of capital generates no cash flows in step 4 at all, and more than \$1 does not increase step 4's cash flows. For notational convenience, we normalize the investor's required rate of return on capital to 0. Given these normalizations, the manager can induce the investor to supply sufficient capital to run the technology by ensuring that the security $\tilde{\psi}$ has an expected payoff to the investor of at least \$1.

With the \$1 capital input, the firm's cash flows (in step 4) are given by the realization of the random variable $\tilde{\theta}$, which assumes one of three possible values: $\theta_l \equiv \mu - e$, or $\theta_m \equiv \mu$ or $\theta_h \equiv \mu + e$, for some $e > 0$. Each of θ_l and θ_h occurs with probability $\frac{\eta}{2} \in \left(0, \frac{1}{2}\right)$, and θ_m occurs with probability $1 - \eta$. We further assume that $\mu - e > 0$ and $\mu > 1$. In short, the firm's realized cash flows are symmetrically distributed around their mean μ; the cash flows are always positive, and investing in the production technology is efficient.

The security $\tilde{\psi}$ yields a payoff that is potentially contingent on the firm's realized cash flows. Given that the firm's cash flows can take on three values, a security $\tilde{\psi}$ can be described completely via the triple (ψ_l, ψ_m, ψ_h), with $\psi_i = \psi(\theta_i)$. The securities we study include debt, equity, hybrids that mix debt and equity, and, more generally, any security $\tilde{\psi}$ that satisfies: (1) limited liability for the manager ($\tilde{\psi} \leq \tilde{\theta}$); (2) limited liability for the investor ($0 \leq \tilde{\psi}$); and (3) monotonicity in the

---

7 We justify this assumption by noting that there are obvious and well-known restrictions and frictions that make short-selling difficult, and these restrictions have become even more onerous with the SEC's September 2008 actions (SEC's "Financial Firm Emergency Order" No. 34-58592; Sept. 18, 2008). Even if we allowed for shorting, the informed trader could make no extra profit by shorting, because any decision to short would reveal her information to the market maker.

8 Dye (1998) uses a similar information representation to model privately informed investors.

9 The key feature of the information structure is that it is "tiered," with the manager always weakly better informed than the informed trader, and the informed trader always weakly better informed than the market maker. While we have assumed that—when the informed trader knows that the manager is withholding information—the informed trader does not do anything about what information the manager is withholding, we could have assumed that the informed trader imperfectly knows what the manager knows. We expect that our main results would continue to hold in the latter case, with few changes. We know, for example, that our results are valid with no changes whatsoever if we assume the informed trader learns that, when the manager has withheld information, the informed trader learns into which element of the partition $\{\{\theta_h\}, \{\theta_m, \theta_l\}\}$ the manager's private information falls.

sense that both $\psi(\theta)$ and $\theta - \psi(\theta)$ are weakly increasing in θ, with $\theta_l - \psi(\theta_l) < \theta_h - \psi(\theta_h)$. In the following, we collectively refer to (1)–(3) as the "limited liability and monotonicity assumptions." Assumptions similar to these natural, standard assumptions have been adopted in many studies of securities' design.[10]

Preliminary Analysis

We start the analysis by observing that there are only three possible equilibrium levels for the volume, X, of the aggregate order flow for the security. This follows from the assumption that, when the investor experiences a liquidity shock, the investor must sell 100 percent of her security. Since the informed trader has an interest in masking the identity of her trades, either she does not participate in the securities market, or else she submits a market order for exactly the same quantity of the security $\tilde{\psi}$ the investor sells (which we normalize to 1).[11] Hence, X's only possible values are -1 (if the investor experiences a liquidity shock and the trader stays out of the market), 0 (if either (a) the investor experiences a liquidity shock and the trader trades or (b) neither the investor experiences a liquidity shock nor does the trader trade), or $+1$ (if the investor does not experience a liquidity shock and the trader trades).

Next, we consider how the security $\tilde{\psi}$ is priced by the market maker. Upon observing the manager's disclosure (if any) and the aggregate order flow X, the market maker posts a price $P(d,X)$ for the security, such that $P(d,X) = E(\tilde{\psi} \mid d, X)$. This price is what the investor receives by selling her security. On average, an investor who faces a liquidity shock must anticipate having to sell her security for less than its expected value because of the presence of the informed trader.[12] So, the *ex ante* value of the security to the investor:

$$I(\tilde{\psi} = d) \equiv vE(P(d,X)|\tilde{q} = 1) + (1 - v)E(\tilde{\psi}), \tag{1}$$

typically will be below the unconditional expected value $E(\tilde{\psi})$ of the security. The difference between $E(\tilde{\psi})$ and $I(\tilde{\psi}, d)$ is the firm's "cost of capital" (CC) in this model:

$$CC(\tilde{\psi},d) \equiv E(\tilde{\psi}) - I(\tilde{\psi},d). \tag{2}$$

The Downside Risk and Upside Potential of Securities

In the following analysis, it is often desirable to characterize a security in terms of how the value of the security changes with the firm's realized cash flows, as in the following definition.

Definition 1: (a) The "downside risk" of security (ψ_l, ψ_m, ψ_h) is given by $\Delta_D = \psi_m - \psi_l$; (b) the "upside potential" of security (ψ_l, ψ_m, ψ_h) is given by $\Delta_U = \psi_h - \psi_m$.

The downside risk of a security is the amount by which the payoff of the security drops as a consequence of having the low state θ_l occur, rather than the medium state, and the upside potential of a security is the amount by which the payoff of the security increases as a consequence of having the high state θ_h occur rather than the medium state θ_m. In view of the relationship $\psi_m = \psi_l + \Delta_D$ and $\psi_h = \psi_l + \Delta_D + \Delta_U$, a security with payoffs (ψ_l, ψ_m, ψ_h) can be described

[10] See, e.g., Harris and Raviv (1990), Innes (1990), Nachman and Noe (1994), DeMarzo and Duffie (1999).

[11] Were the informed trader to submit market orders in quantities different from that of the investor, she would reveal her trades and thereby eliminate her information advantage over the market maker.

[12] This fact establishes, as was asserted earlier, that were the investor not subject to a liquidity shock, the investor would not participate in this market. Instead, she would wait until the firm's cash flows realizes its value and collect the cash-flow-contingent payment as specified by the security.

equivalently via the triple $(\psi_l, \Delta_D, \Delta_U)$. It is straightforward to show that, in terms of this triple, the limited liability and monotonicity assumptions are equivalent to the following three conditions: (1) $0 \leq \Delta_i \leq e$ for $i = D,U$, (2) $\Delta_i \neq 0$ for at least one $i = D,U$, (3) $0 \leq \psi_l \leq \theta_l = \mu - e$.

Debt securities, which take the form $\tilde{\psi} = \min\{F, \tilde{\theta}\}$ for some face value F, play an important role in the following. Definition 2 classifies debt instruments in terms of the triple $(\psi_l, \Delta_D, \Delta_U)$.

Definition 2: A security $(\psi_l, \Delta_D, \Delta_U)$ is "risk-free debt" when it is of the form $(\psi_l, 0, 0)$; it is "investment-grade debt" when it is of the form $(\mu - e, \Delta_D, 0)$ where $\Delta_D > 0$; and it is "speculative (or junk) debt" when it is of the form $(\mu - e, e, \Delta_U)$ where $\Delta_U > 0$.

The classification of debt securities makes clear that the basic difference among different types of debt securities is determined by the states in which default occurs (or, alternatively, the probability of default): a firm never defaults on risk-less debt, it defaults on investment-grade debt only if the low state occurs, and it defaults on speculative debt if anything other than the high state occurs.

Equity securities take the form $(\psi_l, \Delta_D, \Delta_U) = (\alpha(\mu - e), \alpha e, \alpha e)$ for some $\alpha \in (0,1]$; hybrid securities are securities that are neither debt nor equity. Below we show that, when all disclosures are voluntary, both equity securities and hybrid securities are dominated by debt securities, and that equity securities may be optimal when some disclosures are mandatory.

Capital Structure-Dependent Disclosure Policies

Lemma 1 characterizes how the manager's preferred disclosure policy varies with the firm's capital structure/security design choice $\tilde{\psi}$.

Lemma 1: Suppose the manager offers the investor a security $(\psi_l, \Delta_D, \Delta_U)$ that satisfies the limited liability and monotonicity conditions. Then, the possible disclosure policies $d(\cdot)$ the security can induce the manager to adopt are the following: $d(\theta_h) = \theta_h$, $d(\theta_l) = \varnothing$, and:

$$\text{if } pe < \Delta_D - (1-p)\Delta_U, \quad \text{then } d(\theta_m) = \varnothing; \tag{3}$$

$$\text{if } pe > \Delta_D - (1-p)\Delta_U, \quad \text{then } d(\theta_m) = \theta_m; \tag{4}$$

and if each strict inequality in (3) and (4) is replaced by an equality, then the security can induce the manager either to disclose, or to not disclose, $\tilde{\theta} = \theta_m$.

Lemma 1 shows that any security that satisfies the monotonicity and limited liability conditions always induces the manager to disclose θ_h and never induces him to disclose θ_l, but whether the security induces the manager to disclose θ_m depends upon how large the security's upside potential is as compared to its downside risk. The proposition thus formalizes one of the claims made in Section III: a firm's capital structure—that is, the securities it uses to raise capital from investors—affects the firm's disclosure policy.

The intuition for the proposition is as follows. Consider downside risk first and recall that the manager's disclosure decision is based on whether the disclosure increases the market's perceptions of the value of his residual claim $\theta - \psi(\theta)$. It follows that, when the value of his residual claim rises substantially due to having the medium state rather than the low state occur, the manager wants the capital market to know it. Ergo, as the security's downside risk decreases, the manager is more likely to prefer disclosing θ_m. A similar analysis shows that the manager's preference for disclosing θ_m also depends on the upside potential of the security. When the inequality in (3) and (4) holds, we say that security $\tilde{\psi}$ induces the manager to adopt a *limited (expansive) disclosure policy* and also that $d(\cdot)$ is the *disclosure policy induced by* $\tilde{\psi}$.

It is also worth noting that Lemma 1 is consistent with, and extends, some classic comparative statics results of voluntary disclosure. For example, the proposition implies that, as e decreases—i.e., as the volatility of the firm's cash flow decreases (in the sense of second-order stochastic dominance)—then a firm is less likely to engage in expansive disclosure. This implication is consistent with Jung and Kwon's (1988) comparative static that, as a firm's cash flow volatility decreases, the threshold above which it discloses information (and below which it withholds information) increases. Also, by rewriting Lemma 1 so as to emphasize how the switch from limited to expansive disclosure depends on the probability p that the firm receives information, we see that there is a threshold probability $p^* \equiv \frac{\Delta_D - \Delta_U}{e - \Delta_U}$ such that the manager engages in expansive (limited) disclosure when $p > p^*$ $(p < p^*)$. This is consistent with Dye's (1985) and Jung and Kwon's (1988) result that the disclosure threshold declines—and firms are more likely to disclose their private information—as the probability they receive information increases.

Informed Trader's Optimal Trading Strategy

When the manager discloses the information he received, the market maker is at informational parity with the informed trader. Hence, the trader does not trade, since she knows there is no opportunity to earn positive expected trading profits.[13] But, recall that, when the manager discloses nothing, the trader, unlike the market maker, knows whether the nondisclosure is due to the manager not receiving information or due to the manager withholding information. The insider will recognize that the market maker is likely to underprice (overprice) the security when the insider knows the reason the manager made no disclosure is because the manager did not receive (withheld) information, and this likely underpricing (overpricing) will trigger the insider to trade (stay out of the market). This provides the intuition for Lemma 2.

Lemma 2: The informed trader's optimal trading strategy is to submit a market order to buy security $\tilde{\psi}$ when the manager makes no disclosure and the trader knows that the reason the manager made no disclosure was because the manager did not receive information; in all other circumstances, the trader does not trade.[14]

Determinants of a Firm's Cost of Capital

In this section, we provide a detailed analysis of how a firm's cost of capital depends on the design of its securities and the firm's disclosure policy.

For a risk-free security, that is, a security whose payout $\psi(\theta)$ is constant across all realizations of $\tilde{\theta}$, the cost of capital is 0 because the market maker never misprices such a security. Even a risky security—i.e., one whose payout $\psi(\theta)$ varies with θ—is correctly priced by the market maker in three situations: (1) the manager discloses the firm's realized cash flows; (2) the manager makes no disclosure and the aggregate market order for security $\tilde{\psi}$ is $X = 1$ (because $X = 1$ occurs only when the trader places a market order and the investor did not experience a liquidity shock, which reveals that the trader knew the manager's nondisclosure was due to the manager not receiving information); or (3) the manager makes no disclosure and the aggregate market order for security $\tilde{\psi}$ is $X = -1$ (because $X = -1$ occurs only when the investor, and not the trader, places a market order, which reveals that the trader knew the manager's nondisclosure was due to the manager withholding information).

What creates trading losses for an investor who experiences a liquidity shock is the situation

---

[13] We assume that traders who make 0 expected trading profits do not trade. This assumption is justified by any, even trivial, transaction costs from trading.

[14] The proof of Lemma 2 is straightforward and omitted.

in which the manager makes no disclosure and $X = 0$. From the market maker's perspective, this last event is consistent with either: (4) neither the investor nor the informed trader places a market order for security $\tilde{\psi}$, or (5) both the investor (because she faced a liquidity shock) and the market maker (because he knew that the manager's nondisclosure was due to the manager not having received information) place a market order for security $\tilde{\psi}$. In this case, the trader is better informed than the market maker, giving rise to the investor's trading losses. By determining how probable this situation is, and the size of the investor's expected trading losses when it occurs, we can calculate the firm's cost of capital. The next proposition displays the results of this calculation, for any security and any disclosure policy. (In stating Proposition 1, we write $d = 1$ for an expansive disclosure policy (in place of $d(\theta_m) = \theta_m$) and $d = 0$ for a limited disclosure policy (in place of $d(\theta_m) = \varnothing$).)

Proposition 1: The firm's cost of capital when it issues security $\tilde{\psi} = (\psi_l, \Delta_D, \Delta_U)$ and adopts disclosure policy d is given by:

$$CC(\tilde{\psi}, d) = A_U^d \times \Delta_U + A_D^d \times \Delta_D, \tag{5}$$

where

$$A_U^d = v(1-p)\frac{\eta}{2}\frac{p(1-v)\left(\frac{\eta}{2} + (1-d)(1-\eta)\right)}{(1-p)v + p(1-v)\left(\frac{\eta}{2} + (1-d)(1-\eta)\right)};$$

$$A_D^d = v(1-p)\frac{\eta}{2}\frac{p(1-v)\left(\frac{\eta}{2} + d(1-\eta)\right)}{(1-p)v + p(1-v)\left(\frac{\eta}{2} + (1-d)(1-\eta)\right)}.$$

Before discussing Proposition 1, we introduce Corollary 1, which helps to illustrate some of the economic consequences of the proposition.

Corollary 1:

(a) $0 < A_D^0 < A_U^1 < A_U^0 < A_D^1 < \frac{\eta}{2}$; and
(b) $A_D^0 + A_U^0 < A_D^1 + A_U^1 < \frac{\eta}{2}$.

Corollary 1 shows that the firm's cost of capital varies with each of: (1) the security's downside risk Δ_D and its upside potential Δ_U; (2) the specifics of the information and trading environments in which the securities are traded (i.e., the probability that the manager receives information p and the probability that the investor is subject to a liquidity shock v); (3) the technology that produces the firm's cash flows (i.e., the probability distribution of $\tilde{\theta}$, as summarized by the parameter η); and (4) the manager's disclosure policy d.

The ordering of the coefficients $A_U^1 < A_U^0$ shows that the upside potential of a security contributes less to the firm's cost of capital for an expansive disclosure policy than for a limited disclosure policy. Stated differently, *more disclosure favors securities with higher potential.* By contrast, the result $A_D^1 > A_D^0$ in Corollary 1 shows that the downside risk of a security contributes more to the firm's cost of capital when the firm's disclosure policy is expansive ($d = 1$) than when the firm's disclosure policy is limited ($d = 0$). Stated alternatively, *less disclosure favors securities with higher downside risk.*

The intuition for the result is that the firm's cost of capital is the same as the investor's expected trading losses. The investor bears trading losses only when she sells her security following a liquidity shock and the informed trader buys the security. Note that the probability these two events both occur is independent of whether the manager adopts an expansive or limited disclosure policy, as this probability is $v \times (1 - p)$ in all cases.[15] Hence, the only reason the investor's expected trading losses would be higher under one disclosure policy than another is if the equilibrium price at which the investor sells her security when the aggregate order flow is 0 is lower under one disclosure policy than the other. To isolate the effects of expanded disclosure on a security's downside risk, consider a security that has only downside risk (i.e., $\Delta_U = 0$). Such a security provides two distinct possible payoffs to its holder, a "low" payoff when $\tilde{\psi} = \psi_l$ or a "high" payoff (which is the same whether $\tilde{\psi} = \psi_m$ or $\tilde{\psi} = \psi_h$). Under the expansive disclosure policy, the manager only withholds information when the security's payoff is "low," while under the limited disclosure policy, the manager withholds information both when the security's payoff is "low" and also sometimes when the security's payoff is "high" (specifically, when $\tilde{\psi} = \psi_m$). Consequently, when the market maker calculates the equilibrium price when $d = \emptyset$ and $X = 0$ for a security with no upside potential, the calculation will result in a lower price when the manager adopts the expansive disclosure policy than when he adopts the limited disclosure policy. This provides intuition for the ordering $A_D^1 > A_D^0$.

The result $A_U^1 < A_D^0$ in Corollary 1 shows that the upside potential of a security contributes less to the firm's cost of capital for an expansive disclosure policy than for a limited disclosure policy. Stated differently, *more disclosure favors securities with higher upside potential.* The intuition proceeds along lines similar to that in the preceding paragraph.[16]

The results reported in Corollary 1 also show that, if the upside potential and downside risk of a security are the same, then the firm's cost of capital is higher under expansive disclosure than under limited disclosure; that is, with $\Delta \equiv \Delta_D = \Delta_U$, $CC(\tilde{\psi}, 0) = (A_D^0 + A_U^0) \times \Delta < (A_D^1 + A_U^1) \times \Delta = CC(\tilde{\psi}, 1)$.[17]

The orderings depicted in the corollary are important to studying both the design of optimal securities while holding a firm's disclosure policy fixed, and the optimal choice among disclosure policies, as will be seen from the results in Propositions 3, 4, and 5 below.[18]

---

[15] This probability $v(1 - p)$ follows from the fact that v is the probability of a liquidity shock, and the informed trader only buys when she knows that the reason for the manager not making a disclosure is that the manager did not receive any information, which occurs with probability $1 - p$.

[16] To isolate the effects of expanded disclosure on a security's upside potential, consider a security that has only upside potential (so $\Delta_D = 0$). Such a security provides two distinct possible payoffs to its holder, a "low" payoff (which is the same whether $\tilde{\psi} = \psi_l$ or $\tilde{\psi} = \psi_m$) or a "high" payoff when $\tilde{\psi} = \psi_h$. Under the limited disclosure policy, the manager always withholds information when the security's payoff is "low" (i.e., regardless of whether $\tilde{\psi} = \psi_l$ or $\tilde{\psi} = \psi_m$), while under the expansive disclosure policy, the manager sometimes discloses information even though the security's payoff is "low" (specifically, when $\tilde{\psi} = \psi_m$). Consequently, when the market maker calculates the equilibrium price when $d = \emptyset$ and $X = 0$ for a security with no downside risk, the calculation will result in a lower price when the manager adopts the limited disclosure policy than when he adopts the expansive disclosure policy. This provides intuition for the ordering $A_U^1 < A_U^0$.

[17] When the aggregate order flow is 0 and the manager makes no disclosure, under expansive disclosure the security's equilibrium price depends on the manager withholding information only when the security's payoff is ψ_l, whereas under the limited disclosure policy the manager withholds information both when the security's payoff is ψ_l and when its payoff is ψ_m. Since, for a security with equal downside risk and upside potential, $\psi_m = E[\tilde{\psi}] > \psi_l$ (as: $E[\tilde{\psi}] = \psi_m$ for a security with equal downside risk and upside potential), it follows that the firm's equilibrium price will be lower when $d = \emptyset$ and $X = 0$ under the expansive disclosure policy than under the limited disclosure policy, which explains why $CC(\tilde{\psi}, 0) < CC(\tilde{\psi}, 1)$ in that case. This provides intuition for the ordering $A_D^0 + A_U^0 < A_D^1 + A_U^1$.

[18] There are many additional comparative statics that can be derived from Proposition 1. To save space, these comparative statics are not presented here, but they are available from the authors.

III. OPTIMAL SECURITY DESIGN AND DISCLOSURE POLICY

In this section, holding fixed the manager's expansive or limited disclosure policy, we identify the manager's most preferred security among all securities that satisfy the limited liability and monotonicity conditions.

It is intuitive, and it is possible to show formally, that an optimal security minimizes the firm's cost of capital, $A_D^d\Delta_D + A_U^d\Delta_U$, subject to a suitable set of constraints. When the security is designed to induce the manager to implement disclosure policy $d \in \{0,1\}$, the program that identifies the optimal security is the following:

Optimal Security Design d **(OSD**d**) Program**:

$$\min_{\psi_l,\Delta_D,\Delta_U} \quad A_D^d\Delta_D + A_U^d\Delta_U \tag{6}$$

subject to:

$$(1-2d)pe \le (1-2d)(\Delta_D - (1-p)\Delta_U); \tag{7}$$

$$1 \le I = E\tilde{\psi} - A_D^d\Delta_D - A_U^d\Delta_U; \tag{8}$$

$$0 \le \psi_l \le \mu - e; \text{ and} \tag{9}$$

$$0 \le \Delta_i \le e, \quad i \in \{\mu,h\}. \tag{10}$$

The OSDd Program incorporates several results derived above: constraint (7) requires that the manager's adopted disclosure policy d be incentive compatible with the manager's security design choice (Lemma 1); constraint (8) requires that the investor be willing to purchase the security (Lemma 2); and constraints (9) and (10) reflect the monotonicity and limited liability constraints. A security is said to be *feasible* if it satisfies all these constraints.

Clearly, a lower bound on the firm's cost of capital is 0. When this bound can be obtained, we say that it is "first-best." The first-best situation arises if and only if $\Delta_D = \Delta_U = 0$, i.e., the security issued is risk-free debt. The maximum amount of capital the manager can raise by issuing a risk-free security is $\mu - e$. Hence, first-best financing is achieved if and only if the required capital to implement the investment (= 1) is less than $\mu - e$, i.e., if and only if the volatility e in the firm's cash flows does not exceed $\mu - 1$.

The next proposition reports that the risk-free debt security achieves first-best and induces the manager to adopt the expansive disclosure policy.

Proposition 2: There is a security that implements the expansive disclosure policy that yields the first-best outcome if and only if $e \le \mu - 1$. That security must take the form of risk-free debt with face value $F \in [1, \mu - e]$; it always induces expansive disclosure.

Proposition 2 provides the first illustration of the sensitivity of the structure of optimal securities to the size of the firm's cash flow volatility, as measured by the parameter e. This sensitivity will be shown to be a central feature of the analysis that follows. Since Proposition 2 establishes the optimal security for $e \le \mu - 1$, the following analysis focuses on the case of $e > \mu - 1$.

Optimal Securities that Implement Limited Disclosure

This section studies optimal securities that implement the limited disclosure policy. The main result is to establish the existence of a hierarchy of optimal securities, where what constitutes an optimal security varies with the volatility parameter e.

To proceed, we observe that we can set $\psi_l = \mu - e$ in Program OSD0 with no loss in profits to the manager. With this substitution, it is clear that Program OSD0 is a linear program in the two variables Δ_D and Δ_U that can be solved geometrically. To obtain the optimal security, all that has to be done is to search for the lowest iso-cost line for the objective function, i.e., the lowest value k among the lines $A_D^0\Delta_D + A_U^0\Delta_U = k$ that lies within this constraint set. Figure 2 illustrates the linear program that yields the optimal security that implements limited disclosure.

The following proposition summarizes the findings from the geometric investigation of the OSD0 Program in the Appendix.

Proposition 3: The optimal security that implements the limited disclosure policy is:

(1) investment-grade debt with face value:

$$\mu - (1-p)e \quad \text{if } \mu - 1 < e \le \frac{\mu - 1}{1 - \left(1 - \frac{\eta}{2} - A_D^0\right)p} \equiv \underline{e}_{ID}^0;$$

(2) investment-grade debt with face value:

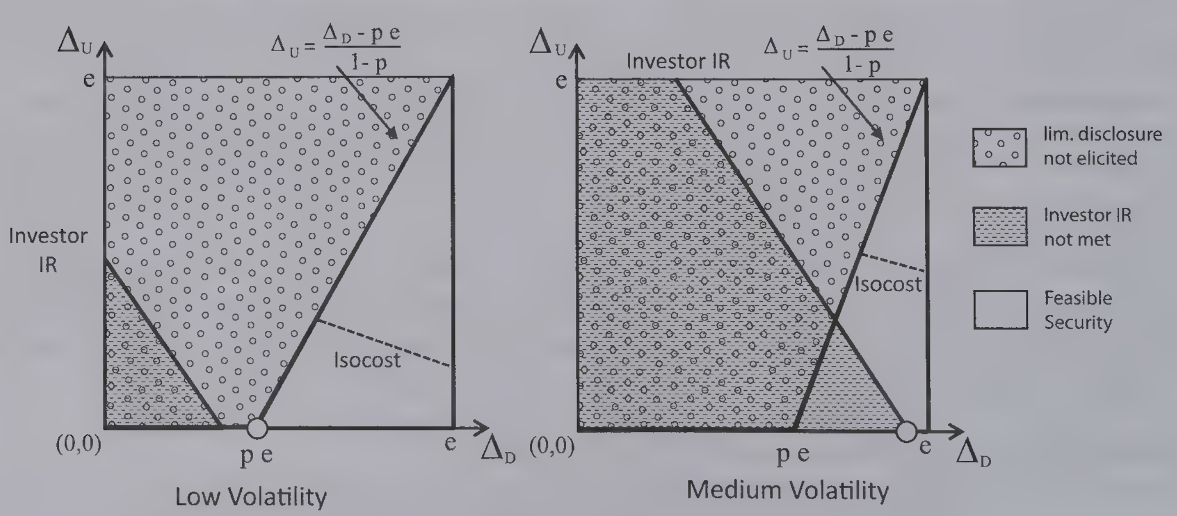
$$\mu - e + \frac{1 + e - \mu}{1 - \frac{\eta}{2} - A_D^0} \quad \text{if } \underline{e}_{ID}^0 < e \le \frac{\mu - 1}{\frac{\eta}{2} + A_D^0} \equiv \underline{e}_{SD}^0;$$

(3) speculative debt with face value:

$$\frac{\mu\left(\frac{\eta}{2} - A_U^0 - 1\right) + 1 + \left(\frac{\eta}{2} + A_D^0\right)e}{\frac{\eta}{2} - A_U^0} \quad \text{if } \underline{e}_{SD}^0 < e < \frac{\mu - 1}{A_D^0 + A_U^0} \equiv \bar{e}_{SD}^0; \quad \text{and}$$

(4) the limited disclosure policy cannot be implemented by any feasible security if $e \ge \bar{e}_{SD}^0$.

FIGURE 2
Optimal Securities that Implement the Limited Disclosure Policy



Proposition 3 establishes that there is a simple two-level hierarchy of optimal securities that induce the manager to implement the limited disclosure policy: investment-grade debt is optimal when the volatility of the firm's cash flows is sufficiently low, and speculative debt is optimal when the volatility of the firm's cash flows is high. When the firm's cash flow volatility is extremely high, no feasible security of any kind implements the limited disclosure policy.

The intuition for the results in Proposition 3 is clear: as noted in the discussion of Proposition 2 above, risk-free debt never induces the manager to implement the limited disclosure policy. Among the risky securities that induce the manager to implement the limited disclosure policy, the manager prefers those securities whose payoffs do not vary much with the firm's cash flows because the absence of much volatility in a security's payoff reduces the informational advantage of the informed trader over the market maker, thereby reducing the trader's expected trading profits and driving down the firm's cost of capital. More than this can be said, however. When choosing among those risky securities that induce the manager to adopt the limited disclosure policy, the manager prefers securities that have only downside risk. Such securities have a lower cost of capital than do securities that also have upside potential, for the reasons described in the discussion following Corollary 1.

When $\psi_l = \mu - e$, securities with only downside risk are of the form $(\psi_l, \Delta_D, \Delta_U) = (\mu - e, \Delta_D, 0)$; such securities are investment-grade debt. An optimal security that implements the limited disclosure policy is, therefore, investment grade debt with downside risk Δ_D that satisfies the disclosure constraint in Equation (7). This explains part (1) of Proposition 3.

As the volatility in the *firm's* cash flows increases, the volatility of the payoff of the optimal security must increase as well. The cost of issuing a security with more volatile payoffs is that the information advantage of the trader over the market maker goes up, which increases the investor's expected trading losses. In order to compensate the investor for her increased trading losses, the manager must issue debt with a higher face value, resulting in a higher cost of capital. This explains part (2) of Proposition 3.

For sufficiently high cash flow volatility levels, even securities with maximal downside risk (but no upside potential) are unable to induce the investor to supply financing. To construct feasible securities in such cases, the manager must resort to securities that have both downside risk and upside potential. Such securities divert all the firm's cash flows to the security holder, unless the most favorable (highest cash flow) state occurs. These securities are speculative debt. This explains part (3) of Proposition 3. Financing the project eventually becomes infeasible at very high levels of the firm's cash flow volatility, because the investor's expected trading losses become so high that no feasible security would compensate her adequately for her expected trading losses. This explains part (4) of Proposition 3.

Optimal Securities that Implement Expansive Disclosure

The program that determines the optimal security that implements the expansive disclosure policy is also a linear program in the two variables Δ_D and Δ_U, which can be solved geometrically. As was the case for the linear program in the preceding section, to obtain the optimal security, all that has to be done is to search for the lowest iso-cost line for the objective function, i.e., the lowest value k among the lines $A_D^1 \Delta_D + A_U^1 \Delta_U = k$ that lies within this constraint set. Figure 3 illustrates the linear program that yields the optimal security that implements expansive disclosure.

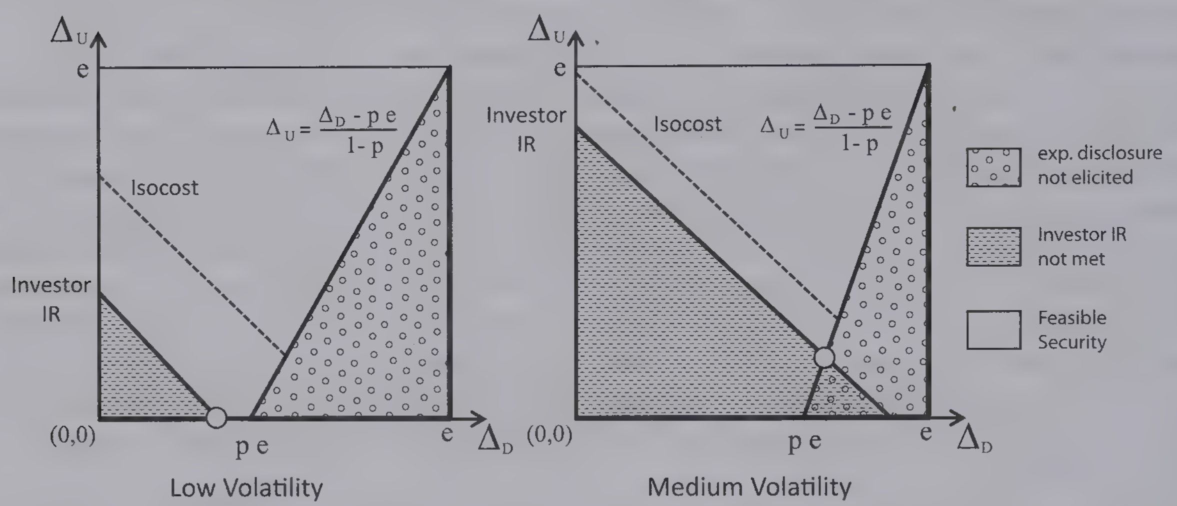
The following proposition summarizes the findings from this geometric investigation.

Proposition 4: An optimal security that implements the expansive disclosure policy is:

(1) risk-free debt with face value:

$$F \in [1, \mu - e] \quad \text{if } e \leq \mu - 1;$$

FIGURE 3
Optimal Securities that Implement the Expansive Disclosure Policy



(2) investment-grade debt with face value:

$$e+\frac{1+e-\mu}{\frac{\eta}{2}-A_U^1}\quad \text{if } \mu-1<e\leq\frac{\mu-1}{1-\left(1-\frac{\eta}{2}-A_D^1\right)p}\equiv \underline{e}_{ID}^1;$$

(3) a hybrid security if:

$$\underline{e}_{ID}^1<e<\frac{\mu-1}{A_D^1+A_U^1}\equiv \bar{e}_{ID}^1;\ \text{and}$$

(4) the expansive disclosure policy cannot be implemented by any feasible security if $e\geq \bar{e}_{ID}^1$.

Proposition 4 establishes that there is a simple three-level hierarchy of optimal securities that induce the manager to implement the expansive disclosure policy: risk-free debt is optimal when the volatility of the firm's cash flows is extremely low; investment-grade debt is optimal when the volatility of the firm's cash flows is too high to make risk-free debt feasible, but is still sufficiently low; and hybrid securities are optimal for higher levels of volatility. As was also true for the limited disclosure policy discussed in the previous section, when the firm's cash flow volatility is extremely high, no feasible security of any kind implements the expansive disclosure policy.

The intuition for part (1) was explained in the discussion of Proposition 2. The intuition for parts (2) and (3) is somewhat similar to that for Proposition 3: as the volatility in the firm's cash flows increases, the volatility of the feasible securities that induce the investor to supply financing also increases. In the discussion following Corollary 1, we observed that the investor's trading losses depend on the price the market maker sets following no disclosure and an order flow of 0. When the manager implements the expansive disclosure policy, the non-disclosure price is a weighted average of ψ_l (in case the manager strategically withheld information) and $E[\tilde{\psi}]=\frac{\eta}{2}\psi_l+(1-\eta)\psi_m+\frac{\eta}{2}\psi_h$ (in case the manager did not receive information). Hence, the non-disclosure price overweights only the low outcome of the security (relative to the prior probability). In order to minimize the investor's trading losses—or, equivalently, minimize the firm's cost of capital—

the manager chooses a security that sets the payoff in the low state to its maximum value, $\psi_l = \mu - e$, but is indifferent between increasing the security's payoff in the intermediate or high state. It follows that debt securities are the optimal securities that induce the expansive disclosure policy (to the extent they satisfy the monotonicity and limited liability constraints). To compensate the investor for her trading losses as the volatility in the firm's cash flows increases, the manager must offer a higher face value on its debt instrument.

However, doing so also reduces the manager's incentives to implement the expansive disclosure policy. When the cash flow volatility extends beyond the critical threshold $\underline{e}_{ID}^{1}$, no feasible investment-grade debt will continue to induce the manager to adopt the expansive disclosure policy. For higher volatilities, the manager needs to issue hybrid securities with not only downside risk, but also upside potential, in order to induce the manager to disclose the intermediate outcome (as required under expansive disclosure policies). Part (4) of Proposition 4 follows for reasons similar to part (5) of Proposition 3: at very high levels of cash flow volatility, the manager cannot finance the project.

Global Hierarchy of Securities and Disclosure Policies

So far, we have established the existence of two hierarchies of securities, one for securities that optimally implement the limited disclosure policy and one for securities that optimally implement the expansive disclosure policy. We next develop a combined hierarchy that involves optimizing across both securities and disclosure policies. We call the optimal security and disclosure policy that solves this joint optimization problem the "globally optimal" security and disclosure policy.

By relying on the results already established, we can get an immediate sense of what constitutes the globally optimal security and disclosure policy in the "corner" case in which the firm's cash flow volatility is sufficiently low (i.e., $e \leq \mu - 1$). For that case, Proposition 2 establishes that the expansive disclosure policy necessarily dominates because risk-free securities induce the expansive disclosure policy but not the limited disclosure policy. Appealing to the continuity of our optimization program, the expansive disclosure policy is also the globally optimal disclosure policy for cash flow volatilities e close to but higher than $\mu - 1$. In that case, the associated globally optimal security is investment-grade debt (by Proposition 4, part 2).

We can also get a sense for what constitutes the globally optimal security and disclosure policy in the other "corner" case in which the firm's cash flow volatility is sufficiently high. When the firm's cash flow volatility is sufficiently high, the only potentially feasible security is one that transfers nearly all of the firm's cash flows to the investor—i.e., one for which both the downside risk and upside potential of the security are nearly maximal: $\Delta_D \approx e \approx \Delta_U$. Given this observation, then the result in Corollary 1 that $A_D^0 + A_U^0 < A_D^1 + A_U^1$ leads immediately to the conclusion that the globally optimal disclosure policy must be the limited disclosure policy, as that result implies that a firm's cost of capital under the limited disclosure policy, $A_D^0\Delta_D + A_U^0\Delta_U \approx (A_D^0 + A_U^0)e$, will be less than its cost of capital under the expansive disclosure policy, $A_D^1\Delta_D + A_U^1\Delta_U \approx (A_D^1 + A_U^1)e$. Appealing to continuity a second time, we can further conclude that the limited disclosure policy is the globally optimal disclosure policy for all cash flow volatilities close to but lower than the maximum volatility for which the project can be implemented. The associated globally optimal security is speculative debt (by Proposition 3).

Having settled the question of what constitutes the globally optimal security and disclosure policy when the firm's cash flow volatility falls in the "corners," it remains to determine the globally optimal security and disclosure policy when the cash flow volatility assumes "intermediate" values. The next proposition provides a complete answer to this question.

Proposition 5: There exists a unique value, e^*, for the volatility of the firm's cash flows, with $e^* \in (\mu - 1, \underline{e}_{ID}^1)$, such that:

(1) if $0 < e \leq e^*$, then expansive disclosure combined with either riskless debt (when $e \leq \mu - 1$) or investment-grade debt (when $e > \mu - 1$), as described in Proposition 4, is optimal;
(2) if $e^* < e < \vec{e}_{SD}^0$, then limited disclosure combined with either investment-grade debt (when $e \leq \underline{e}_{SD}^0$) or speculative debt (when $e > \underline{e}_{SD}^0$), as described in Proposition 3, is optimal;[19] and
(3) if $e \geq \vec{e}_{SD}^0$, then the project cannot be implemented by any feasible security.

Proposition 5 establishes that the manager prefers the expansive disclosure policy in all environments characterized by a sufficiently low level of cash flow volatility, whereas he prefers the limited disclosure policy in all environments characterized by higher levels of cash flow volatility, up to some maximum cash flow volatility beyond which the project cannot be implemented by any disclosure policy (either limited or expansive). In addition, Proposition 5 establishes that, for all cash flow volatilities for which the manager can implement the project, the optimal security is always a debt security. The debt security that is globally optimal is riskless debt when the cash flow volatility is sufficiently low, becoming investment-grade debt as the cash flow volatility increases, and speculative debt as the cash flow volatility increases still further.

Another feature of Proposition 5 worth mentioning is that hybrid securities are never optimal. Hybrid securities have one function: to induce expansive disclosure for parameter values such that standard debt alone would not induce expansive disclosure. Relative to the standard debt security, the hybrid security exhibits more upside potential and less downside risk. Therefore, the manager trades off limited-disclosure downside risk (when using standard debt) and expansive-disclosure upside potential (when using the hybrid). However, following the intuition given in Corollary 1, the limited-disclosure downside risk is less costly than the expansive-disclosure upside potential. It then follows that using the standard debt instrument is preferable to encouraging expansive disclosure using the hybrid security.

Perhaps the most important prediction that emerges from Proposition 5 is the following corollary:

Corollary 2: In equilibrium, there is a negative association between firms' cost of capital and the amount of information voluntarily disclosed by firms.

This prediction follows from two previously established facts. First, as a firm's cash flow volatility e increases, its maximum expected profits decline due to the firm's cost of capital increasing (this follows by combining Propositions 3, 4, and 5). Second, as a firm's cash flow volatility increases, it optimally switches from the expansive disclosure policy to the limited disclosure policy (by Proposition 5).

This negative correlation does not imply, however, that more expansive voluntary disclosure causes firms' cost of capital to decline. On the contrary, the model predicts that, whenever the firm issues risky debt, more disclosure leads to a strict increase in the firm's cost of capital. To see this, first consider investment-grade debt. The cost of capital associated with issuing investment-grade debt, $(\mu - e, \Delta_D, 0)$, is $A_D^d \Delta_D$. Expansive disclosure yields a greater cost of capital for investment-grade debt than does limited disclosure because, by Corollary 1, $A_D^0 < A_D^1$. Next consider speculative debt. The cost of capital associated with issuing speculative debt and adopting the limited disclosure policy is $A_D^0 e + A_D^0 \Delta_U$. This is always less than the cost of capital associated with

[19] The proof of the proposition establishes that $\underline{e}_{ID}^1 < \underline{e}_{ID}^0$. So, since $\underline{e}_{ID}^0 < \underline{e}_{SD}^0$ by Proposition 3, each of the intervals $(e^*, \underline{e}_{SD}^0]$ and $(\underline{e}_{SD}^0, \vec{e}_{SD}^0]$ is nonempty.

issuing speculative debt and adopting the expansive disclosure policy, $A_D^1 e + A_U^1 \Delta_U$ because the downside risk, Δ_D, is higher than the upside potential, Δ_U. Thus, even though for any given risky debt security, there is a positive association between a firm's cost of capital and how much it discloses, when the firm chooses its securities optimally, the equilibrium (hence, predicted) association between a firm's cost of capital and how much the firm discloses is negative.

IV. INTRODUCING MANDATORY DISCLOSURES

In this section, we extend the model by considering how mandatory disclosure requirements affect the manager's choice among securities, the manager's propensity to make supplementary voluntary disclosures, and the firm's cost of capital. We consider two distinct forms of mandatory disclosure requirements. One form, which we label "loss recognition," requires that the manager disclose the low outcome (θ_l) when he knows it. The other form, which we label "gains recognition," requires the manager to disclose the medium outcome (θ_m) when he knows it.

Optimal Security Design under Mandatory Loss Recognition

The next proposition describes the manager's preferred voluntary disclosure policy under mandatory loss recognition. Under mandatory loss recognition, the manager always discloses both the low and the high outcomes. Whether he discloses the medium outcome depends on the relative magnitude of the investor's downside risk and upside potential, as the following proposition indicates:

Proposition 6: Under mandatory loss recognition disclosures, if the manager offers the investor a security $(\psi_l, \Delta_D, \Delta_U)$ that satisfies the limited liability and monotonicity conditions, then the possible voluntary disclosure policies $d(\cdot)$ the security can induce the manager to adopt are the following: $d(\theta_l) = \theta_l$, $d(\theta_h) = \theta_h$, and:

$$\text{if } \Delta_D > \Delta_U, \quad \text{then } d(\theta_m) = \varnothing; \tag{11}$$

$$\text{if } \Delta_D < \Delta_U, \quad \text{then } d(\theta_m) = \theta_m; \tag{12}$$

and if each strict inequality in (11) and (12) is replaced by an equality, then the security can induce the manager either to disclosure, or not to disclose, $\tilde{\theta} = \theta_m$.

The intuition for this result is as follows. Whether the manager discloses the medium outcome depends on how the value of his residual claim compares when he discloses θ_m versus when he does not make a disclosure. Since the manager discloses both the low and high outcome under mandatory loss recognition, the value of his residual claim given the medium outcome is higher (lower) when the manager is exposed to more (less) downside risk than upside potential. Equivalently, the manager chooses to disclose (withhold) the medium outcome when the investor captures less (more) of the downside risk than the upside potential, i.e., $\Delta_D < \Delta_U$ ($\Delta_D > \Delta_U$).

Juxtaposing this disclosure policy with Lemma 1 provides a dramatic illustration of how mandatory disclosures can affect a firm's voluntary disclosure policy. Because the inequality (3) characterizing nondisclosure in Lemma 1 can be written alternatively as $\Delta_U < \frac{\Delta_D - pe}{1-p}$, and since, for all $\Delta_D \in [0, e)$, it is clear that $\frac{\Delta_D - pe}{1-p} < \Delta_D$, it follows that any security $(\psi_l, \Delta_D, \Delta_U)$ that induces the manager not to disclose θ_m when there are no disclosure requirements will also induce the manager not to disclose θ_m when the mandatory loss recognition disclosure requirements are imposed. Thus, in the case of mandatory loss recognition disclosures, holding the security the firm issues fixed, mandatory disclosure inhibits voluntary disclosure. We show in Proposition 7 below, however, that if the manager is allowed to optimize his choice among securities (when the dis-

closure regime changes from all disclosures being voluntary to the mandatory loss recognition regime), then mandatory disclosure leads to expanded voluntary disclosure. This conclusion thus provides yet another illustration of the importance of endogenizing a firm's capital structure when making predictions about disclosures: the predictions about disclosures that emerge from holding capital structure fixed are the opposite of the predictions that emerge when the firm's capital structure is endogenized.

When $\Delta_D < \Delta_U$, the manager engages in full disclosure of his private information and the market maker and the investor are at informational parity with the informed trader. Accordingly, the security will always be correctly priced, the investor will not incur any expected trading losses, the firm's cost of capital will be 0, and the first-best outcome will be achieved. The next proposition shows that the manager can exploit this result and achieve first-best by issuing an equity security.

Proposition 7: Under mandatory loss recognition disclosures, for all $e \in (0,\mu)$, the equity security $\tilde{\psi} = \frac{\tilde{\theta}}{\mu}$ (i.e., $(\psi_l, \Delta_D, \Delta_U) = \left(1 - \frac{e}{\mu}, \frac{e}{\mu}, \frac{e}{\mu}\right)$) induces the manager to engage in full disclosure and achieves the first-best outcome.[20]

Notice that no risky debt security will induce the manager to fully disclose his private information, because no risky debt security satisfies inequality (12). Thus, no risky debt security induces the manager to engage in a policy of full disclosure; hence, no risky debt security completely eliminates the informed trader's information advantage over the market maker and yields the first-best outcome.

We conclude this subsection with two further comments concerning Proposition 7. First, the proposition illustrates that what are considered to be optimal securities can change with the introduction of mandatory disclosure requirements, because equity securities are typically not optimal when all disclosures are voluntary. Second, the proposition illustrates how mandatory disclosures can lead to an increase in the manager's maximum expected profits. We shall see in the next subsection, however, that mandatory disclosures do not always increase the manager's maximum expected profits.

Optimal Security Design under Mandatory Gains Recognition

In this subsection, we discuss the mandatory disclosure regime referred to as "gains recognition" above, which entails that the manager must disclose the medium outcome (θ_m) when he knows it.

Under gains recognition, the manager will always keep silent when he learns that $\tilde{\theta} = \theta_l$ occurs—as he cannot possibly benefit by revealing that the firm's cash flows are the lowest possible level. Because the manager will always disclose $\tilde{\theta} = \theta_h$ when he learns it, and he necessarily discloses $\tilde{\theta} = \theta_m$ when he learns it to be in conformity with the gains recognition mandatory disclosure requirement, it follows that the disclosure policy the manager adopts in the presence of mandatory gains recognition is tantamount to the manager implementing the expansive disclosure policy. From this observation, it is immediate that mandatory gains recognition disclosures sometimes reduces the manager's maximum expected profits relative to the situation in which all disclosures are voluntary. These remarks prove:

Proposition 8: Under mandatory gains recognition disclosures, the disclosure policy the manager is induced to adopt is always the expansive disclosure policy, and

[20] The proof of Proposition is immediate: since $\Delta_D = \Delta_U$, condition (12) holds (weakly) for the equity security $(\psi_l, \Delta_D, \Delta_U) = \left(1 - \frac{e}{\mu}, \frac{e}{\mu}, \frac{e}{\mu}\right)$, so the manager (weakly) prefers full disclosure.

the manager's maximum expected profits are always weakly (in fact, strictly for sufficiently high levels of cash flow volatility) below what the manager could achieve when all disclosures are voluntary.

Viewing mandatory disclosures as a commitment to a disclosure policy, this result shows that commitment to a disclosure policy can reduce the manager's maximum expected profits. While it is commonplace to observe that mandatory disclosure requirements may be undesirable because they generate proprietary costs of disclosure, Propositions 7 and 8 combined indicate that unqualified statements about the desirability or undesirability of expanded mandatory disclosure requirements, or unqualified statements about the desirability of committing to expanded disclosures, typically are invalid even absent concerns about disclosing proprietary information.

V. CONCLUSION

The owners of a firm use all of the instruments at their disposal, including their operating choices, financing choices, investment choices, and disclosure choices, to maximize the expected value of their residual stakes in their firm. Because the array of such instruments available to a firm's owners is in practice so large, any model must inevitably put some restrictions on the range of endogenous variables the owners of a firm are posited to select. The present model demonstrates that, for purposes of studying a firm's disclosure policy, it is undesirable to take the firm's financing decisions as given, because as the firm's financing decisions change, the residual stake over which the owners of the firm have claim change, and the latter change alters the owners' preferred voluntary disclosure policy. We show that the reverse is also true. Namely, the capital structure the owners of a firm prefer also depends on the firm's disclosure policy. These interdependencies imply that, in equilibrium, a firm's capital structure and disclosure policy are jointly determined, and together determine the firm's cost of capital.

APPENDIX

Proof of Lemma 1

In view of the assumed monotonicity of $\theta - \psi(\theta)$, the unique optimality of $d(\theta_h) = \theta_h$ and $d(\theta_l) = \varnothing$ is clear. Consequently, the proof focuses on the specification of $d(\theta_m)$. If Equation (3) describes the manager's disclosure policy, then $E[\tilde{\theta} - \psi(\tilde{\theta}) \mid d]$ yields:

$$E[\tilde{\theta} - \psi(\tilde{\theta})|d = \theta_h] = \theta_h - \psi_h$$

and, by Bayes' rule:

$$E[\tilde{\theta} - \psi(\tilde{\theta})|d = \varnothing] = \frac{(1-p)(\theta_m - E[\tilde{\psi}]) + p\left((1-\eta)(\theta_m - \psi_m) + \frac{\eta}{2}(\theta_l - \psi_l)\right)}{1 - p + p\left(1 - \frac{\eta}{2}\right)}$$

$$= \theta_m - E[\tilde{\psi}] - \frac{p\left((1-\eta)(\psi_m - E[\tilde{\psi}]) + \frac{\eta}{2}(e - E[\tilde{\psi}] + \psi_l)\right)}{1 - p\frac{\eta}{2}}$$

where:

$$E[\tilde{\psi}] = (1-\eta)\psi_m + \frac{\eta}{2}(\psi_l + \psi_h) = \psi_m + \frac{\eta}{2}[(\psi_h - \psi_m) - (\psi_m - \psi_l)].$$

Given the restriction to truth-telling, were the manager to disclose $d = \theta_m$, then necessarily:

$$E[\tilde{\theta} - \psi(\tilde{\theta}) | d = \theta_m] = \theta_m - \psi_m.$$

So, for the disclosure policy in Equation (3) to be an equilibrium, we require:

$$E[\tilde{\theta} - \psi(\tilde{\theta}) | d = \theta_m] \leq E[\tilde{\theta} - \psi(\tilde{\theta}) | d = \varnothing],$$

i.e.:

$$\theta_m - \psi_m \leq \mu - E[\tilde{\psi}] - \frac{p\left((1-\eta)(\psi_m - E[\tilde{\psi}]) + \frac{\eta}{2}(e - E[\tilde{\psi}] + \psi_l)\right)}{1 - p\frac{\eta}{2}}$$

which simplifies to:

$$pe \leq \Delta_D - (1-p)\Delta_U.$$

The proof of Equation (4) is analogous and omitted.

Proof of Proposition 1

The investor's expected payoff is:

$$\begin{aligned} I &= (1-\upsilon)E[\tilde{\psi}] + \upsilon\begin{pmatrix} \Pr(d=\varnothing, X=-1|q=1)P(\varnothing,-1) \\ +\Pr(d=\varnothing, X=0|q=1)P(\varnothing,0) \\ +\Pr(d=\theta_h)\psi_h + \Pr(d=\mu)\psi_m \end{pmatrix} \\ &= E[\tilde{\psi}] + \upsilon\begin{pmatrix} \Pr(d=\varnothing, X=-1|q=1)P(\varnothing,-1) \\ +\Pr(d=\varnothing, X=0|q=1)P(\varnothing,0) \\ +\Pr(d=\theta_h)\psi_h + \Pr(d=\mu)\psi_m - E[\tilde{\psi}] \end{pmatrix} \end{aligned}$$

Note that:

$$\Pr(d=\varnothing, X=-1|q=1) + \Pr(d=\varnothing, X=0|q=1) + \Pr(d=\theta_h) + \Pr(d=\mu) = 1.$$

The price $P(\varnothing, X)$ can be written as:

$$\begin{aligned} P(\varnothing, X) &= \psi_m + \Pr(\theta_h|\varnothing, X, d)\Delta_U - \Pr(\theta_l|\varnothing, X, d)\Delta_D \\ &= E[\tilde{\psi}] - \left(\psi_m + \frac{\eta}{2}[\Delta_U - \Delta_D]\right) + \psi_m + \Pr(\theta_h|\varnothing, X, d)\Delta_U - \Pr(\theta_l|\varnothing, X, d)\Delta_D \\ &= E[\tilde{\psi}] + \left(\Pr(\theta_h|\varnothing, X, d) - \frac{\eta}{2}\right)\Delta_U - \left(\Pr(\theta_l|\varnothing, X, d) - \frac{\eta}{2}\right)\Delta_D. \end{aligned}$$

Then, I can be written as:

$$I = E[\tilde{\psi}] + v\left(\begin{array}{c} \Pr(d=\varnothing, X=-1|q=1,d)\left(\begin{array}{c} E[\tilde{\psi}] + \left(\Pr(\theta_h|\varnothing, X=-1,d) - \frac{\eta}{2}\right)\Delta_U \\ -\left(\Pr(\theta_l|\varnothing, X=-1,d) - \frac{\eta}{2}\right)\Delta_D \end{array}\right) \\ +\Pr(d=\varnothing, X=0|q=1,d)\left(\begin{array}{c} E[\tilde{\psi}] + \left(\Pr(\theta_h|\varnothing, X=0,d) - \frac{\eta}{2}\right)\Delta_U \\ -\left(\Pr(\theta_l|\varnothing, X=0,d) - \frac{\eta}{2}\right)\Delta_D \end{array}\right) \\ +\Pr(d=\theta_h)\left(E[\tilde{\psi}] + \left(1-\frac{\eta}{2}\right)\Delta_U + \frac{\eta}{2}\Delta_D\right) \\ +\Pr(d=\mu)\left(E[\tilde{\psi}] - \frac{\eta}{2}(\Delta_U - \Delta_D)\right) - E[\tilde{\psi}] \end{array}\right) = E[\tilde{\psi}]$$

$$+ v\left(\begin{array}{c} \Pr(d=\varnothing, X=-1|q=1,d)\left(\begin{array}{c} \left(\Pr(\theta_h|\varnothing, X=-1,d) - \frac{\eta}{2}\right)\Delta_U \\ -\left(\Pr(\theta_l|\varnothing, X=-1,d) - \frac{\eta}{2}\right)\Delta_D \end{array}\right) \\ +\Pr(d=\varnothing, X=0|q=1,d)\left(\begin{array}{c} \left(\Pr(\theta_h|\varnothing, X=0,d) - \frac{\eta}{2}\right)\Delta_U \\ -\left(\Pr(\theta_l|\varnothing, X=0,d) - \frac{\eta}{2}\right)\Delta_D \end{array}\right) \\ +\Pr(d=\theta_h)\left(1-\frac{\eta}{2}\right)\Delta_U + \frac{\eta}{2}\Delta_D \\ -\Pr(d=\mu)\frac{\eta}{2}(\Delta_U - \Delta_D) \end{array}\right)$$

and therefore:

$$A_U^d = v\left(\begin{array}{c} \Pr(d=\varnothing, X=-1|q=1,d)\left(\frac{\eta}{2} - \Pr(\theta_h|\varnothing, X=-1,d)\right) \\ +\Pr(d=\varnothing, X=0|q=1,d)\left(\frac{\eta}{2} - \Pr(\theta_h|\varnothing, X=0,d)\right) \\ -\left(1-\frac{\eta}{2}\right)\Pr(d=\theta_h) + \frac{\eta}{2}\Pr(d=\mu) \end{array}\right)$$

$$A_D^d = v\left(\begin{array}{c} \Pr(d=\varnothing, X=-1|q=1,d)\left(\Pr(\theta_l|\varnothing, X=-1,d) - \frac{\eta}{2}\right) \\ +\Pr(d=\varnothing, X=0|q=1,d)\left(\Pr(\theta_l|\varnothing, X=0,d) - \frac{\eta}{2}\right) \\ -\frac{\eta}{2}\Pr(d=\theta_h) - \frac{\eta}{2}\Pr(d=\mu) \end{array}\right).$$

Market Maker's Beliefs if the Manager Implements the Limited Disclosure Policy

Let $s_{knowcash} = \theta$ indicate that the manager learns that cash flows will be θ, and $s_{knowcash} = \varnothing$ indicate that the manager receives no information about cash flows. Suppose the manager does not disclose the median value and the market maker conjectures that the trader does not buy the security if she sees $s_{withheld} = 1$ and purchases the security if she sees $s_{withheld} = 0$. Then, the market maker's beliefs are as follows.

If $X = 1$ and $d = \varnothing$, then the trader must have purchased the security because $s_{withheld} = 0$. Hence, the market maker beliefs equal the prior beliefs.

$$\Pr(\theta_h, |\varnothing, X = 1, d = 0) = \Pr(\theta_h | s_{withheld} = 0) = \frac{\eta}{2}$$

$$\Pr(\theta_l | \varnothing, X = 1, d = 0) = \Pr(\theta_l | s_{withheld} = 0) = \frac{\eta}{2}.$$

If $X = 0$ and $d = \varnothing$, then either (1) the trader purchased the security because $s_{withheld} = 0$ and there was a liquidity shock (which occurs with probability $(1 - p)v$), or (2) the trader did not purchase the security because $s_{withheld} = 1$ (which implies $s_{knowcash} \in \{\mu, \theta_l\}$) and there was no liquidity shock (which occurs with probability $p\left(1 - \frac{\eta}{2}\right)(1 - v)$). Hence, the market maker's beliefs are:

$$\Pr(\theta_h | \varnothing, X = 0, d = 0) = \frac{\Pr(\theta_h, d = \varnothing, X = 0)}{\Pr(d = \varnothing, X = 0)} = \frac{\frac{\eta}{2}(1 - p)v}{(1 - p)v + \left(1 - \frac{\eta}{2}\right)p(1 - v)} < \frac{\eta}{2}$$

$$\Pr(\theta_l | \varnothing, X = 0, d = 0) = \frac{\Pr(\theta_l, d = \varnothing, X = 0)}{\Pr(d = \varnothing, X = 0)} = \frac{\frac{\eta}{2}((1 - p)v + p(1 - v))}{(1 - p)v + \left(1 - \frac{\eta}{2}\right)p(1 - v)} > \frac{\eta}{2}.$$

If $X = -1$ and $d = \varnothing$, then from $X = -1$, the market maker can infer that the trader did not purchase the security because $s_{withheld} = 1$. Hence:

$$\Pr(\theta_h | \varnothing, X = -1, d = 0) = 0$$

$$\Pr(\theta_l | \varnothing, X = -1, d = 0) = \frac{\frac{\eta}{2}}{1 - \eta + \frac{\eta}{2}} = \frac{\eta}{2 - \eta}.$$

Based on these beliefs and:

$$\Pr(d=\varnothing, X=-1|q=1, d=0) = \left(1-\frac{\eta}{2}\right)p$$

$$\Pr(d=\varnothing, X=0|q=1, d=0) = (1-p)$$

$$\Pr(d=\theta_h) = \frac{\eta}{2}p$$

$$\Pr(d=\mu) = 0$$

we can compute:

$$A_U^0 = v\left(\begin{array}{c} \Pr(d=\varnothing, X=-1|q=1, d=0)\left(\frac{\eta}{2}-\Pr(\theta_h|\varnothing, X=-1, d=0)\right) \\ +\Pr(d=\varnothing, X=0|q=1, d=0)\left(\frac{\eta}{2}-\Pr(\theta_h|\varnothing, X=0, d=0)\right) \\ -\left(1-\frac{\eta}{2}\right)\Pr(d=\theta_h)+\frac{\eta}{2}\Pr(d=\mu) \end{array}\right)$$

$$= v\left(\left(1-\frac{\eta}{2}\right)p\left(\frac{\eta}{2}-0\right)+(1-p)\left(\frac{\eta}{2}-\frac{\frac{\eta}{2}(1-p)v}{(1-p)v+\left(1-\frac{\eta}{2}\right)p(1-v)}\right)-\left(1-\frac{\eta}{2}\right)\frac{\eta}{2}p\right)$$

$$= v(1-p)\frac{\eta}{2}\frac{\left(1-\frac{\eta}{2}\right)p(1-v)}{(1-p)v+\left(1-\frac{\eta}{2}\right)p(1-v)}$$

and

$$A_D^0 = v\left(\begin{array}{c} \Pr(d=\varnothing, X=-1|q=1, d=0)\left(\Pr(\theta_l|\varnothing, X=-1, d=0)-\frac{\eta}{2}\right) \\ +\Pr(d=\varnothing, X=0|q=1, d=0)\left(\Pr(\theta_l|\varnothing, X=0, d=0)-\frac{\eta}{2}\right) \\ -\frac{\eta}{2}\Pr(d=\theta_h)-\frac{\eta}{2}\Pr(d=\mu) \end{array}\right)$$

$$= v\left(\left(1-\frac{\eta}{2}\right)p\left(\frac{\eta}{2-\eta}-\frac{\eta}{2}\right)+(1-p)\left(\frac{\frac{\eta}{2}((1-p)v+p(1-v))}{(1-p)v+\left(1-\frac{\eta}{2}\right)p(1-v)}-\frac{\eta}{2}\right)-\frac{\eta}{2}\frac{\eta}{2}p\right)$$

$$= v(1-p)\frac{\eta}{2}\frac{\frac{\eta}{2}p(1-v)}{(1-p)v+\left(1-\frac{\eta}{2}\right)p(1-v)}.$$

Market Maker's Beliefs if the Manager Implements the Expansive Disclosure Policy

Following the same logic as before, we can calculate:

$$\Pr(\theta_h|\varnothing,X=0,d=1)=\frac{\Pr(\theta_h,d=\varnothing,X=0)}{\Pr(d=\varnothing,X=0)}=\frac{\frac{\eta}{2}(1-p)v}{(1-p)v+\frac{\eta}{2}p(1-v)}<\frac{\eta}{2}.$$

$$\Pr(\theta_l|\varnothing,X=0,d=1)=\frac{\Pr(\theta_l,d=\varnothing,X=0)}{\Pr(d=\varnothing,X=0)}=\frac{\frac{\eta}{2}((1-p)v+p(1-v))}{(1-p)v+\frac{\eta}{2}p(1-v)}>\frac{\eta}{2}.$$

If $X = -1$ and $d = \varnothing$, then from $X = -1$, the market maker can infer that the trader did not purchase the security because $s_{withheld}=1$ and $s_{knowcash}=\mu$. Hence:

$$\Pr(\theta_h|\varnothing,X=-1,d=1)=0$$

$$\Pr(\theta_l|\varnothing,X=-1,d=1)=1.$$

Based on these beliefs and:

$$\Pr(d=\varnothing,X=-1|q=1,d=1)=\frac{\eta}{2}p$$

$$\Pr(d=\varnothing,X=0|q=1,d=1)=(1-p)$$

$$\Pr(d=\theta_h)=\frac{\eta}{2}p$$

$$\Pr(d=\mu)=(1-\eta)p$$

we can compute:

$$A_U^1=v(1-p)\frac{\eta}{2}\frac{\frac{\eta}{2}p(1-v)}{(1-p)v+\frac{\eta}{2}p(1-v)}$$

and:

$$A_D^1=v(1-p)\frac{\eta}{2}\frac{\left(1-\frac{\eta}{2}\right)p(1-v)}{(1-p)v+\frac{\eta}{2}p(1-v)}.$$

■

Proof of Proposition 3

The Optimal Security Design $d = 0$ (OSD0) Program

$$\min_{\psi_l, \Delta_D, \Delta_U} A_D^0 \Delta_D + A_U^0 \Delta_U \tag{13}$$

subject to:

$$pe \leq \Delta_D - (1-p)\Delta_U; \tag{14}$$

$$1 + e - \mu \leq \left(1 - \frac{\eta}{2} - A_D^0\right)\Delta_D + \left(\frac{\eta}{2} - A_U^0\right)\Delta_U; \tag{15}$$

$$0 \leq \Delta_i \leq e; \quad i \in \{D, U\}. \tag{16}$$

The feasible set of Δ_D and Δ_U is then determined by Equation (15) to be that subset of the square $[0,e] \times [0,e]$ in (Δ_D, Δ_U)-space lying below the line:

$$\Delta_U = \frac{\Delta_D - pe}{1-p} \tag{17}$$

and above the investor's "supply-financing" line:

$$\Delta_U = \frac{1 + e - \mu - \left(1 - \frac{\eta}{2} - A_D^0\right)\Delta_D}{\frac{\eta}{2} - A_U^0}. \tag{18}$$

We start by observing that, in all non-first best cases (i.e., when $1 + e - \mu > 0$), the investor's "supply financing" line (18) intersects the positive orthant in (Δ_D, Δ_U)-space. The key to determining the smallest value attained by the objective function (13) in the constraint set is the observation that the (downward-sloping) iso-cost lines of the objective function (13) are flatter than the slope of the line (18), as the following lemma states.

Lemma 3

$$\frac{A_D^0}{A_U^0} < \frac{1 - \eta/2 - A_D^0}{\eta/2 - A_U^0}.$$

Proof

Rewrite

$$A_D^0 \text{ and } A_U^0 \text{ as } A_D^0 = \frac{\eta}{2}A^0 \text{ and } A_U^0 = \left(1 - \frac{\eta}{2}\right)A^0,$$

where:

$$A^0 = v(1-p)\frac{\eta}{2}\frac{p(1-v)}{(1-p)v + \left(1 - \frac{\eta}{2}\right)p(1-v)}.$$

With this notation:

$$\frac{A_D^0}{A_U^0} = \frac{\frac{\eta}{2}A^0}{\left(1-\frac{\eta}{2}\right)A^0} = \frac{\frac{\eta}{2}}{1-\frac{\eta}{2}} \tag{19}$$

and $\frac{1-\eta/2-A_D^0}{\eta/2-A_U^0} = \frac{1-\frac{\eta}{2}-\frac{\eta}{2}A^0}{\frac{\eta}{2}-\left(1-\frac{\eta}{2}\right)A^0}$. Then it is easy to confirm that the inequality in the statement of the lemma is equivalent to $0 < 1 - \eta$, which always holds. ■

Given this fact, it follows that the optimal security is defined by the "most southwest" point of the constraint set (14)–(16), as illustrated in Figure 2. We identify what that "most southwest" point is by studying the following four mutually exclusive and exhaustive cases (for $e > \mu - 1$).

- In case 1 ("low volatility"), the point on the line (18) lying on the Δ_D-axis, i.e., the point $(\Delta_D^0, \Delta_U^0) \equiv \left(\frac{1+e-\mu}{1-\frac{\eta}{2}-A_D^0}, 0\right)$, lies to the left of $(\Delta_D, \Delta_U) = (pe, 0)$.[21] This case occurs when $\left(\frac{1+e-\mu}{1-\frac{\eta}{2}-A_D^0} \leq pe\right)$ i.e., when $e \leq \frac{\mu-1}{1-\left(1-\frac{\eta}{2}-A_D^0\right)p} \equiv \underline{e}_{ID}^0$. In this case, the "southwest-most" (Δ_D, Δ_U) in the feasible set is $(pe, 0)$. This corresponds to the security $(\psi_l, \Delta_D, \Delta_U) = (\mu - e, pe, 0)$. This security is investment-grade debt with face value $\mu - (1 - p)e$.
- In case 2 ("medium volatility"), (Δ_D^0, Δ_U^0) lies to the right of $(\Delta_D, \Delta_U) = (pe, 0)$ but to the left of $(\Delta_D, \Delta_U) = (e, 0)$. This case occurs when $e > \underline{e}_{ID}^0$ and $\frac{1+e-\mu}{1-\frac{\eta}{2}-A_D^0} \leq e$, i.e., when $\underline{e}_{ID}^0 < e \leq \frac{\mu-1}{\frac{\eta}{2}+A_D^0} \equiv \underline{e}_{SD}^0$. In this case, the "southwest-most" (Δ_D, Δ_U) in the feasible set is (Δ_D^0, Δ_U^0). This corresponds to the security $(\psi_l, \Delta_D, \Delta_U) = \left(\mu - e, \frac{1+e-\mu}{1-\frac{\eta}{2}-A_D^0}, 0\right)$. This security is investment-grade debt with face value $\mu - e + \frac{1+e-\mu}{1-\frac{\eta}{2}-A_D^0}$.
- In case 3 ("high volatility"), (Δ_D^0, Δ_U^0) lies to the right of $(\Delta_D^0, \Delta_U) = (e, 0)$ and the point (Δ_D, Δ_U) on the line (18) evaluated at $\Delta_D = e$ lies inside the box $[0,e] \times [0,e]$. This case occurs when $e > \underline{e}_{SD}^0$ and $\frac{1+e-\mu-\left(1-\frac{\eta}{2}-A_D^0\right)e}{\left(\frac{\eta}{2}-A_U^0\right)} \leq e$, so $\underline{e}_{SD}^0 < e < \frac{\mu-1}{A_D^0+A_U^0} \equiv \vec{e}_{SD}^{\,0}$. In this case, the "southwest-most" (Δ_D, Δ_U) in the feasible set is $\left(e, \frac{1+e-\mu-\left(1-\frac{\eta}{2}-A_D^0\right)e}{\frac{\eta}{2}-A_U^0}\right)$. This corresponds to the security $(\psi_l, \Delta_D, \Delta_U) = \left(\mu - e, e, \frac{1+e-\mu-\left(1-\frac{\eta}{2}-A_D^0\right)e}{\frac{\eta}{2}-A_U^0}\right)$. This security is speculative debt with face value $\frac{\mu\left(\frac{\eta}{2}-A_U^0-1\right)+1+\left(\frac{\eta}{2}+A_D^0\right)e}{\frac{\eta}{2}-A_U^0}$.
- In case 4 ("extremely high volatility"), (Δ_D^0, Δ_U^0) lies to the right of $(\Delta_D, \Delta_U) = (e, 0)$, but the point (Δ_D, Δ_U) on the line (18) evaluated at $\Delta_D = e$ lies outside the box $[0,e] \times [0,e]$. This case occurs when $e \geq \vec{e}_{SD}^{\,0}$. In this case, no feasible security can implement the limited disclosure policy.

Substituting the values of the downside risk and upside potential of the securities that have been identified to be optimal in each of these four cases into the expression for the firm's cost of capital in Proposition 1 completes the proof of Proposition 3. ■

---

[21] Also note that $(\Delta_\mu^0, \Delta_h^0)$ lies to the right of $(\Delta_\mu, \Delta_h) = (0, 0)$ whenever first-best is not attainable.

Proof of Proposition 4

This proof is similar in style to Proposition 3, so its proof is omitted. ■

Proof of Proposition 5

We start the proof of the proposition by proving the following Lemma.

Lemma 4: For any $e > \underset{\sim}{e}_{ID}^1$ for which there exists a feasible security—call it "security 1"—that implements the expansive disclosure policy, there exists a feasible debt security that implements the limited disclosure policy that increases the manager's expected profits relative to that obtained from security 1.

Proof: Suppose $e > \underset{\sim}{e}_{ID}^1$ and some security, say "security 1," eliciting expansive disclosure exists. Then, we are in case 3 of Proposition 4. As the proof of Proposition 4 shows, optimal securities in this case are hybrid securities determined by those securities on the "supply-financing" line:

$$\Delta U = \frac{1 + e - \mu - \left(1 - \frac{\eta}{2} - A_D^1\right)\Delta_D}{\frac{\eta}{2} - A_U^1}$$

that lie in the constraint set for Program OSD1. Single out the optimal security on this line that also lies on the boundary line:

$$pe = \Delta_D - (1-p)\Delta_U \tag{20}$$

of the constraint set. Call this latter security "security 2," and denote its components by $(\psi_l^2, \Delta_D^2, \Delta_U^2)$. Security 2 necessarily generates at least as high expected profits for the manager as does security 1, because security 2 is optimal for Program OSD1 and security 1 is merely feasible for this program. Recall by Lemma 1 that securities that lie on the line (20) can induce the manager to adopt either the limited or expansive disclosure policy. In particular, this is true for security 2.

We now make two claims: security 2 is feasible for Program OSD0, and security 2 yields the manager higher expected profits under that program than under Program OSD1. This will prove that the manager is better off implementing the limited disclosure policy upon being given security 2 rather than the expansive disclosure policy. Notice that the firm's cost of capital when the manager implements the expansive disclosure policy with security 2 satisfies:

$$\begin{aligned} A_D^1\Delta_D + A_U^1\Delta_U &= A_D^1(pe + (1-p)\Delta_U) + A_U^1\Delta_U \\ &= (A_D^1 + A_U^1)\Delta_U + A_D^1 p(e - \Delta_U) > (A_D^0 + A_U^0)\Delta_U + A_D^0 p(e - \Delta_U) \\ &= A_D^0(pe + (1-p)\Delta_U) + A_U^0\Delta_U = A_D^0\Delta_D + A_U^0\Delta_U, \end{aligned} \tag{21}$$

where the inequality follows from Corollary 1 and the first and last equalities follow from security 2 lying on (20). Of course (21) is equivalent to:

$$\left(1 - \frac{\eta}{2} - A_D^1\right)\Delta_D + \left(\frac{\eta}{2} - A_U^1\right)\Delta_U \le \left(1 - \frac{\eta}{2} - A_D^0\right)\Delta_D + \left(\frac{\eta}{2} - A_U^0\right)\Delta_U. \tag{22}$$

(22), along with the feasibility of $(\psi_l^2, \Delta_D^2, \Delta_U^2)$ for Program 0SD1 and the fact that $(\psi_l^2, \Delta_D^2, \Delta_U^2)$ lies on (20), implies the feasibility of $(\psi_l^2, \Delta_D^2, \Delta_U^2)$ for Program OSD0. (21) implies that the manager's expected profits are strictly higher by implementing the limited disclosure policy than the expansive disclosure policy upon being given security 2. This proves the two claims.

To conclude the proof, recall that Proposition 3 established that debt securities always optimally implement the limited disclosure policy, so there is some debt security, say "security 3," that is at least weakly preferable to security 2 that implements the limited disclosure policy. Since security 2 is weakly preferable to security 1, the proof of the lemma is complete. ■

Proof of Proposition 5

The proof of this proposition consists of the following three observations. First, the function defining the manager's maximum expected profits when he implements the limited disclosure policy is linear in e over the interval $e \in (\mu - 1, \underset{\sim}{e}^1_{ID}]$; the same is also true of the function defining the manager's maximum expected profits when he implements the expansive disclosure policy (for this same interval $(\mu - 1, \underset{\sim}{e}^1_{ID}]$.[22] Second, the manager's maximum expected profits are strictly higher under the expansive disclosure policy than under the limited disclosure policy at, and—by continuity—in the vicinity of, $e = \mu - 1$. (This follows from the discussion surrounding Proposition 2.) Third, the manager's maximum expected profits are strictly higher under the limited disclosure policy than under the expansive disclosure policy at, and—by continuity—in the vicinity of $e = \underset{\sim}{e}^1_{ID}$. (This follows directly from Lemma 4 above.) Hence, these expected profit functions must cross exactly once in the interval $(\mu - 1, \underset{\sim}{e}^1_{ID}]$, which defines the point e_S in the statement of the proposition. The conclusion about the global optimality of risk-free debt, investment-grade debt, or speculative debt then follows directly from Propositions 3 and 4.

Proof of Proposition 7

Disclosure of θ_m is preferred over non-disclosure of θ_m when the following condition is satisfied:

$$\theta_m - \psi_m \geq E[\tilde{\theta} - \tilde{\psi} | d = \varnothing, d(\theta_m) = \varnothing],$$

where RHS of the inequality is given by:

$$E[\tilde{\theta} - \tilde{\psi} | d = \varnothing, d(\theta_m) = \varnothing] = \frac{p(1-\eta)(\theta_m - \psi_m) + (1-p)\left(\theta_m - \left(\frac{\eta}{2}\psi_h + (1-\eta)\psi_m + \frac{\eta}{2}\psi_l\right)\right)}{p(1-\eta) + (1-p)}$$

Simplification yields:

$$\theta_m - \psi_m \geq E[\tilde{\theta} - \tilde{\psi} | d = \varnothing, d(\theta_m) = \varnothing]$$

$$\theta_m - \psi_m \geq \theta_m - \frac{\psi_h + \psi_l}{2}$$

$$\frac{\psi_h + \psi_l}{2} \geq \psi_m$$

[22] Because: (1) by Proposition 3, the manager's expected profit function when he adopts the limited disclosure policy is linear in e for e in the interval $(\mu - 1, \underset{\sim}{e}^0_{ID}]$; (2) by Proposition 4, the manager's expected profit function is linear in e for e in the interval $(\mu - 1, \bar{e}^1_{ID}]$, and so, since $\underset{\sim}{e}^1_{ID} < \bar{e}^1_{ID}$, *a fortiori* is linear over the interval $(\mu - 1, \underset{\sim}{e}^1_{ID}]$; (3) $\underset{\sim}{e}^1_{ID} = \frac{\mu-1}{1-(1-\frac{\eta}{2}-A^1_D)p} < \frac{\mu-1}{1-(1-\frac{\eta}{2}-A^0_D)p} = \underset{\sim}{e}^0_{ID}$, because $A^0_D < A^1_D$, by Corollary 1 above; and so (4) both of these expected profit functions are linear in e for e in the interval $(\mu - 1, \underset{\sim}{e}^1_{ID}]$.

$\Delta_U \geq \Delta_D$.

The case for which non-disclosure of θ_m is preferred over disclosure of θ_m is identical.

Next, we derive the trader's optimal trading strategy as a function of the manager's disclosure policy. Suppose the manager discloses both the low and high state but not the intermediate state. Then, the trader prefers to buy the security when $s_{withheld} = 0$ if and only if her expected purchasing price, $\upsilon P(\varnothing,0) + (1 - \upsilon)P(\varnothing,1)$, is less than the expected payoff of the security, $E[\tilde{\psi}]$. $P(\varnothing,1)$ equals $E[\tilde{\psi}]$ (because the market maker can infer that $s_{withheld} = 0$) and $P(\varnothing,0)$ is a weighted average of $E[\tilde{\psi}]$ and ψ_m. Hence, $y(\varnothing,0) = 1$ is optimal if $E[\tilde{\psi}] > \psi_m$ or $\Delta_U > \Delta_D$. This condition is incompatible with the manager choosing not to disclose θ_m. Hence, the trader does not buy the security when $s_{withheld} = 0$. The trader prefers to buy the security when $s_{withheld} = 1$ if and only if her expected purchasing price, $\upsilon P(\varnothing,0) + (1 - \upsilon)P(\varnothing,1)$, is less than the payoff of the security, ψ_m. $P(\varnothing,1)$ equals ψ_m (because the market maker can infer that $s_{withheld} = 1$) and $P(\varnothing,0)$ is a weighted average of $E[\tilde{\psi}]$ and ψ_m. Hence, $y(\varnothing,1) = 1$ is optimal if $\psi_m > E[\tilde{\psi}]$ or $\Delta_U < \Delta_D$. This condition is consistent with the manager choosing not to disclose θ_m. Hence, the trader buys the security when $s_{withheld} = 1$. Clearly, the trader never buys the security when the manager fully discloses his private information. ■

REFERENCES

Bias, B., T. Mariotti, G. Plantin, and J.-C. Rochet. 2007. Dynamic security design: Convergence to continuous time and asset pricing implications. *Review of Economic Studies* 74 (2): 345–390.

Botosan, C. A. 1997. Disclosure level and the cost of equity capital. *The Accounting Review* 72 (3): 323–349.

———, and M. A. Plumlee. 2002. Re-examination of disclosure level and the expected cost of equity capital. *Journal of Accounting Research* 40 (1): 21–40.

Casamatta, C. 2003. Financing and Advising: Optimal financial contracts with venture capitalists. *Journal of Finance* 58 (5): 2059–2086.

Christensen, P., G. Feltham, and M. Wu. 2002. Cost of capital in residual income for performance evaluation. *The Accounting Review* 77 (1): 1–23.

Demange, G., and G. Laroque. 1995. Private information and the design of securities. *Journal of Economic Theory* 65 (1): 233–257.

DeMarzo, P., and D. Duffie. 1999. A liquidity-based model of security design. *Econometrica* 67 (1): 65–99.

Dye, R. A. 1985. Disclosure of nonproprietary information. *Journal of Accounting Research* 23 (1): 123–145.

———. 1998. Investor sophistication and voluntary disclosures. *Review of Accounting Studies* 3 (3): 261–287.

Fulghieri, P., and D. Lukin. 2001. Information production, dilution costs, and optimal security design. *Journal of Financial Economics* 61 (1): 3–42.

Harris, M., and A. Raviv. 1990. Capital structure and the informational role of debt. *The Journal of Finance* 45 (2): 321–349.

———, and ———. 1991. The theory of capital structure. *The Journal of Finance* 46 (1): 297–356.

Innes, R. 1990. Limited liability and incentive contracting with ex-ante action choices. *Journal of Economic Theory* 52: 45–67.

Jorgensen, B., and M. Kirschenheiter. 2003. Discretionary risk disclosures. *The Accounting Review* 78 (2): 449–469.

Jung, W.-O., and Y. K. Kwon. 1988. Disclosure when the market is unsure of information endowment of managers. *Journal of Accounting Research* 26 (1): 146–153.

Kanodia, C. S., A. Mukherji, H. Sapra, and R. Venugopalan. 2000. Hedge disclosures, futures prices and production distortions. *Journal of Accounting Research* 38 (Supplement): 53–82.

Korajczyk, R. A., D. J. Lucas, and R. L. McDonald. 1991. The effect of information releases on the pricing and timing of equity issues. *Review of Financial Studies* 4 (4): 685–708.

———, ———, and ———. 1992. Equity issues with time-varying asymmetric information. *Journal of Finan-*

cial and Quantitative Analysis 27 (3): 397–417.

Kyle, A. 1985. Continuous auctions and insider trading. *Econometrica* 53 (6): 1315–1335.

Myers, S., and N. S. Majluf. 1984. Corporate financing and investment decisions when firms have information that investors do not have. *Journal of Financial Economics* 13 (2): 187–221.

Nachman, D. C., and T. H. Noe. 1994. Optimal design of securities under asymmetric information. *Review of Financial Studies* 7 (1): 1–44.

Pae, S. 2002. Discretionary disclosure, efficiency, and signal informativeness. *Journal of Accounting and Economics* 33 (3): 279–311.

Rahi, R. 1995. Partially revealing rational expectations equilibria with nominal assets. *Journal of Mathematical Economics* 24 (2): 137–146.

Repullo, R., and J. Suarez. 2004. Venture capital finance: A security design approach. *Review of Finance* 8 (1): 75–108.

Stocken, P. C. 2000. Credibility of voluntary disclosure. *The RAND Journal of Economics* 31 (2): 359–374.

Sunder, J. 2006. Information spillovers and capital structure. Working paper, Northwestern University.

THE ACCOUNTING REVIEW
Vol. 86, No. 3
2011
pp. 887–914

American Accounting Association
DOI: 10.2308/accr.00000038

Dividends, Share Repurchases, and Tax Clienteles: Evidence from the 2003 Reductions in Shareholder Taxes

Jennifer L. Blouin
University of Pennsylvania

Jana S. Raedy
The University of North Carolina at Chapel Hill

Douglas A. Shackelford
The University of North Carolina at Chapel Hill and National Bureau of Economic Research

ABSTRACT: This study jointly evaluates firm-level changes in investor composition and shareholder distributions following a 2003 reduction in the dividend and capital gains tax rates for individuals. We find that directors and officers, but not other individual investors, rebalanced their portfolios to maximize after-tax returns in light of the new tax rules. We also find that firms adjusted their distribution policy (specifically, dividends versus share repurchases) in a manner consistent with the altered tax incentives for individual investors. To our knowledge, this is the first study to employ simultaneous equations to estimate both shareholder and managerial responses to the 2003 rate reductions. We find that the generalized method of moments (GMM) estimates are substantially stronger than OLS estimates, consistent with our expectation that investor and manager responses are simultaneously determined. Failure to estimate systems of equations may account for some of the weak and conflicting results from prior studies of the 2003 rate reductions.

Keywords: *shareholder taxes; dividends; share repurchases; tax clienteles.*

Data Availability: *Data used in this study are available from public sources identified in the study.*

The authors have benefited greatly from comments from Dan Dhaliwal and two anonymous referees, discussions with Jeff Brown, Raj Chetty, Emmanuel Saez, and Scott Weisbenner, and research assistance from Kevin Markle. We are responsible for all remaining errors.

Editor's note: Accepted by Steven Kachelmeier, with thanks to Dan Dhaliwal for serving as editor on a previous version.

Submitted: November 2007
Accepted: September 2010
Published Online: April 2011

I. INTRODUCTION

This study employs simultaneous equations to assess the responses of both investors and companies to the 2003 reductions in dividend and capital gains tax rates. We predict that individual investors, the only ones affected by the reduction in shareholder taxes, rebalanced their portfolios to maximize after-tax returns in light of the new tax rules. We also predict that firms adjusted their distribution policies (specifically, dividends versus share repurchases) to maximize share value—i.e., distributing profits in a manner that was most attractive to their investors after considering shareholder taxes. With regard to investor responses, we find evidence that insiders (i.e., directors and officers) increased holdings in their own companies if their dividend-repurchase mix reflected the new tax incentives. However, we find no evidence that other individual investors rebalanced their portfolios. With regard to managerial responses, we find that firms with disproportionately large individual holdings modified their payouts in a manner consistent with the altered tax incentives. However, changes in dividend and repurchase policy were not immediate; firms deferred widespread, substantial changes until the second quarter following enactment.

To our knowledge, this is the first study to jointly estimate investor and firm responses to changes in shareholder taxes. The fact that both investors and firms can change their behavior following a change in shareholder taxes presents an identification problem. To illustrate, suppose we test for an association between dividend yields and individual stock ownership and find that the correlation becomes more positive following a reduction in individual dividend tax rates. Such a finding is consistent with both (1) individuals switching to high-dividend-paying firms following the tax cut (i.e., a tax clientele response) and (2) firms that are held mostly by individuals increasing their dividends following the tax reduction (a firm payout response). To distinguish between investors rebalancing their portfolios and firms altering their distributions, this study estimates simultaneous equations. Prior studies have focused on either investors or managers, but not both, and results have been weak and in conflict. By adopting simultaneous equations, we find evidence of substantially stronger investor and managerial responses to a decline in shareholder tax rates.

To maximize the power of our tests, we compare a firm's ownership and dividend-repurchase mix before and after the largest change in U.S. dividend taxation. The Jobs and Growth Tax Relief Reconciliation Act of 2003 (JGTRRA) reduced, for individuals only, the maximum, statutory dividend tax rate from 38.6 percent to 15 percent. It also lowered the maximum, statutory individual capital gains tax rate, which applies to share repurchases, from 20 percent to 15 percent. In many ways, JGTRRA is an ideal legislative setting for testing an association between shareholder taxes and payout policy. The scope of JGTRRA was narrow. Its genesis was individual dividend tax reduction. Its primary amendment was individual capital gains tax relief.[1] The final bill did little more than reduce dividend and capital gains tax rates. Nonetheless, the economic effects were huge.[2] Because of its narrow focus and big impact, the JGTRRA provides a much stronger setting than other shareholder tax rate changes, which were either much smaller (e.g., Tax Relief Act of 1997) or involved widespread overhaul of the tax system that affected far more than just shareholder taxes (e.g., Tax Reform Act of 1986).

The JGTRRA dividend and capital gains tax cuts should have altered the optimal mix of dividends and repurchases for at least some individual investors. As a result, we expect that some individuals rebalanced their portfolios so that a higher proportion of their equity returns came in

---

1 For the legislative history, see Auerbach and Hassett (2007).

2 The size of the tax savings has been very large. In September 2007, the U.S. Congress' Joint Committee on Taxation reported that the 2003 dividend and capital gains tax rates will cost the Treasury $632 billion between 2007 and 2011, the largest tax expenditure in the tax code.

the form of dividends. Consistent with such portfolio rebalancing, Desai and Dharmapala (2011) report that total U.S. equity investments shifted from foreign countries, whose companies did not qualify for the lower dividend tax rates, to foreign countries, whose companies did quality for the lower dividend tax rates. However, to our knowledge, no one has documented whether individual investors rebalanced their much larger holdings of domestic holdings in a manner consistent with the changed tax incentives following passage of the JGTRRA. This study addresses that void by studying the domestic portfolio rebalancing of three different individual investor groups.

To attract individual investors who were looking for more dividends following passage of the legislation, we expect that at least some firms increased the dividend portion of their distributions. Consistent with such a managerial response, Chetty and Saez (2005) document that dividend initiations jumped in 2003. Brown et al. (2007) find that reductions in share repurchases funded these 2003 dividend initiations, with such substitution limited to companies where directors and officers held disproportionately large shares. They find no similar managerial responses among the set of firms that account for almost all dividend issuances—i.e., the firms paying dividends before JGTRRA, or among firms without high insider ownership.[3] Comparing executive compensation in 2003 and 2002, Aboody and Kasznik (2008) also reach mixed conclusions about the changes in the dividend-repurchase mix.

We conduct a more comprehensive study of managerial responses to JGTRRA by looking at dividend-paying firms (which far exceed initiators in number and payout), studying both insiders and other individual investors, and extending the analysis to include changes in distribution policies through 2005 (two years beyond the extant JGTRRA studies). Lengthening the investigation period enables us to calibrate how long it took investors and managers to respond to JGTRRA. We also estimate a system of equations and find that results are statistically and economically more significant under simultaneous equations than under ordinary least-squares.

To test for investor and firm responses, we compare the percentage of shares held before and after JGTRRA for three groups of individual investors: insiders (i.e., directors and officers of the firm), other non-executive individuals investing on their own account, and mutual fund investors.[4] We aggregate each firm's dividends and repurchases during the eight quarters immediately preceding the quarter of enactment and compute the ratio of dividends to total payout (dividends plus repurchases).[5] We compare that ratio to the one formed with aggregated dividends and repurchases for the eight quarters immediately after the quarter of enactment. We then test for an association between the change in shareholders and change in the dividend-repurchase mix. This difference-in-differences approach mitigates the likelihood of spurious conclusions arising from omitted correlated variables. Estimating a system of equations enables us to determine whether the association is driven by tax clientele effects, payout changes, or both.

We find evidence consistent with both investor and firm responses to JGTRRA. However, insiders are the only investors who appear to have altered their holdings in response to JGTRRA. We find stronger evidence that firms modified their payout policy in response to changes in

---

3 Brown et al. (2007) are careful to state that their inferences about substitution are limited to 2003 dividend initiators. That said, because their sample includes both initiators and non-initiating firms, it was unclear to us whether the findings reported in Table 6 of their study shed any light on the firm responses of non-initiators. Private conversations with the authors confirmed that their conclusions were limited to 2003 dividend initiators. In addition, to enable us to investigate the non-initiators in their study more closely, the authors kindly provided us with their data, for which we were most appreciative. We replicated their results, and, consistent with the inferences in their study and our conversations with them, we found that their results only hold for firms that initiated in 2003.

4 Many mutual fund investors are not individuals and the earnings for many mutual funds held by individuals are not subject to JGTRRA, e.g., 401(k) investments. Unfortunately, we cannot observe the extent to which JGTRRA affects a specific mutual funds' investor base. This limitation biases against our finding a response by mutual fund investors.

5 We aggregate to reduce some of the noise arising because repurchases, unlike regular quarterly dividends, are uneven and irregular.

shareholder taxes. In particular, we find that the movement toward distributing a larger proportion of profits as dividends was greatest among not only those companies held disproportionately by individual investors, particularly directors and officers, but also those firms where other individuals and mutual funds had large holdings. We find that firms began to substantially alter their distribution policies in the second quarter following the passage of JGTRRA, consistent with firms needing time to adjust their dividend and repurchase policies.

We find that the economic significance of the 2003 legislation is substantially greater using simultaneous equations than using ordinary least-squares (OLS). This is an important contribution of this study because, to our knowledge, we are the first study to jointly evaluate shareholder and firm behavior. Based solely on our OLS results, we would erroneously conclude that no investors rebalanced their portfolios and that the firm responses were limited to those companies with heavy insider ownership. Using generalized method of moments (GMM) estimation of the system of equations, which is the appropriate approach given our expectation that investors and managers simultaneously readjusted their behavior after 2003, we find that some investors, namely individuals, altered their portfolios and that firms were responsive to insiders, mutual funds, and other individual tax considerations.

The next section develops testable hypotheses. Section III details the research design. Section IV presents the empirical findings. Closing remarks follow in Section V.

II. HYPOTHESIS DEVELOPMENT

Framework

This study does not attempt to tackle the longstanding puzzle of why firms pay dividends when they could distribute profits through share repurchases, which remain tax-advantaged, though less so, even after JGTRRA.[6] Rather, in a nutshell, we: (1) accept the fact that investors desired and some firms paid dividends before JGTRRA (obviously for non-tax reasons), (2) assume that the mix of dividends and repurchases was optimal before passage of the legislation, (3) expect that the large tax rate reductions for dividends (compared with the relatively modest reductions for capital gains) led some individuals to rebalance their portfolios in favor of dividend income, and (4) predict that some firms, in response to the changing tax incentives for individual investors, increased the portion of their distributions in the form of dividends after JGTRRA. The remainder of this section elaborates on these relations to develop formal hypotheses about the impact of JGTRRA on shareholder distributions.

To develop the intuition for our hypotheses, we start with a simple framework.[7] Absent taxes, suppose that all investors hold optimally diversified portfolios of the risk-free asset and the market portfolio. In that setting, shareholder ownership of stocks should not vary across investors with similar risk preferences. Now suppose some investors (call them individuals) become taxed on their dividend income. These individual investors will no longer hold the original, pre-tax, optimally diversified portfolio. Rather, they will underweight their portfolio in tax-disfavored, dividend-paying stocks and overweight their portfolio in tax-favored, no-dividend stocks. This shift will boost the price of no-dividend stocks and drive down the price of dividend-paying

[6] For a sampling of the "dividend puzzle" literature, see Miller and Modigliani (1961), Feldstein and Green (1983), Bagwell and Shoven (1989), Berhheim (1991), DeAngelo et al. (2006), Jagannathan et al. (2000), Fama and French (2001), Grullon and Michaely (2002), Dhaliwal and Li (2006), Gordon and Dietz (2008), and Chetty and Saez (2010), among many others. On a different note, even though both repurchases and dividends now face a maximum tax rate of 15 percent, repurchases remain tax-advantaged for at least two reasons. First, sellers can offset the tax basis of the shares that they sell against the proceeds from the sale in computing their capital gains. Second, they can offset those capital gains with capital losses that otherwise might not be deductible. This contrasts with dividends, where the entire amount is taxed upon receipt.

[7] We appreciate the contributions of an anonymous referee in sketching out this framework.

stocks. These price movements will entice non-individual investors (call them tax-exempts) to hold more dividend stocks and less no-dividend stocks. As a result, each investor ultimately would hold the portfolio that features his optimal trade-off between risk and after-tax return; the share price of each stock would equate supply and demand; and a heterogeneous mix of shareholders would emerge endogenously. One example of an empirical study of such tax clientele responses is Dhaliwal et al. (1999), who document increases in institutional holdings when firms initiate dividends, which are tax-disadvantaged, to individual investors. [8]

Similarly, if individuals were taxed more heavily on dividends than on share repurchases, then they would overweight their portfolios with stocks that distributed disproportionate amounts of profits through repurchases, as compared to dividends.[9] Meanwhile, non-individual investors would overweight their portfolios with stocks that distributed disproportionate amounts of profits through dividends, as compared to share repurchases. In this setting, managers would choose both the level and mix (dividends versus share repurchases) of shareholder distributions that maximize the firm's stock price. Their distribution policy would affect both the equilibrium price and the equilibrium mix of investors and changes in their payout policy will induce a change in the mix of investors that own the firm.[10]

Now consider a reduction in the dividend tax rate. Individual investors would rebalance their portfolios, adding more dividend-paying stocks than was optimal under the prior high-dividend tax regime. This will drive up the price of dividend-paying stocks, altering the optimal portfolio mix for tax-exempt investors who would now shift from dividend-paying stocks to no-dividend stocks at the margin. At the same time that investors would be adjusting their portfolios in light of the changes in dividend tax policy, managers would be revising their level and mix of shareholder distributions in light of the new tax policy to continue to maximize their stock price. As a result, both the investor mix (individuals versus tax-exempts) and the firm's distribution policy (dividends versus share repurchases) would change after a reduction in the dividend tax rate.

Determining the relative importance of the investor response and the firm's response would require joint evaluation of both investors' and managers' incentives. Studies that examine only individual investors' new tax-motivated demand for dividends might erroneously attribute all of the increase in dividends to a clientele effect. In the extreme, this is true even if investors did not rebalance their portfolio, but rather, firms simply increased their dividend payouts. Likewise, studies (such as several prior examinations of JGTRRA) that focused solely on firm's new tax-driven supply of dividends might erroneously attribute increased dividend income by individuals to a payout response when the result was actually due to individual investors rebalancing their portfolios. In the tests below, we find evidence of both a clientele and a managerial response to shareholder tax rate reductions.

Investor Responses

We begin our hypothesis development by focusing on potential investor responses to JGTRRA. We assume that because the decline in the dividend tax rate (from 38.6 percent to 15

---

[8] Moser and Puckett (2009) include the JGTRRA in their study of whether there is a positive association between the portion of dividend-paying securities in tax-advantaged institutions' portfolios and the dividend penalty. Although they find evidence of a positive association, their analysis is confounded by the classification of financial institutions as tax-advantaged institutions (Blouin 2009).

[9] We are assuming that tax-sensitive investors do not interpret any potential dividend and repurchase signals differently from other market participants.

[10] For a formal equilibrium model of stock ownership and stock price, see Guenther and Sansing (2010). Motivated by recent empirical findings of a relation between tax-exempt ownership and dividend tax capitalization, they model and provide empirical evidence that dividend tax capitalization reflects all investors' weighted average tax rates, where the weighting varies with the investors' risk tolerances. An important difference between their study and ours is that they focus on prices and ownership, whereas we focus on payouts and ownership.

percent) exceeded the decrease in the capital gains tax rate (from 20 percent to 15 percent), the net effect of JGTRRA was to make firms that distributed profits mostly through dividends more attractive for individual investors than those that distribute profits mostly through share repurchases. Assuming investors were holding the optimal portfolio (considering risk and taxes) before JGTRRA's enactment, we predict that, after enactment, individual investors altered their portfolios to receive a higher percentage of their returns in the form of dividends.

> **H1:** Individual investors responded to JGTRRA by increasing their holdings in stocks that distribute larger proportions of their profits through dividends.

H1 notwithstanding, it is important to recognize that investors cannot freely rebalance their portfolios. Aside from the commissions and other transaction costs that investors face on all trades, taxable investors pay capital gains taxes on any excess of the proceeds from the sale of the stock over the basis of the stock. Thus, some investors may accept an inferior portfolio, rather than incur the tax and non-tax costs of rebalancing their portfolio. This is one reason why we might not observe investors engaging in widespread rebalancing following JGTRRA.

Firm Responses

We anticipate that firms will respond to individual investors' enhanced interest in dividends by increasing the portion of profits that they distribute in the form of dividends. Those firms wishing to retain or increase their holdings by individual investors are likely to distribute more of their profits as dividends, following the enactment of JGTRRA, than other firms. Unfortunately, we cannot observe the pool of future shareholders that firms hope to attract by adjusting their payout policy. Therefore, we look to their shareholder mix at passage and assume that firms with greater individual ownership at enactment would be more likely to alter their distribution policy to distribute more of their profits as dividends than those firms with less individual ownership. To the extent a firm's current investor mix is not a good predictor of managers' desired investor mix in the future, our tests are biased against finding a firm response to JGTRRA.

> **H2:** Managers responded to JGTRRA by distributing a larger portion of their profits as dividends. The extent to which managers increased their dividend percentage increased with the individual ownership of their firm.

The tension in H2 arises from several reasons why managers might not have adjusted their dividend-repurchase mix in response to JGTRRA. First, the JGTRRA tax rate reductions were scheduled to expire in five years at the end of 2008 (later deferred to the end of 2010), and 2004 Democratic Presidential candidate, John Kerry, pledged to restore the higher dividend tax rates for the two highest tax brackets, if elected. Since dividends tend to be sticky and the markets historically punish firms for decreasing dividends, many firms may have chosen to leave their distribution policy unchanged, delay any change until after the 2004 elections, or turn to one-time special dividends (which we ignore in this study). Second, dividends are purported to play an important role by alleviating asymmetric information through conveying private information to the market. Tax-motivated dividend changes might undermine this signal. Third, large increases in dividends could adversely affect the firm's compensation structure, particularly to the extent that the firm relies on stock options, which are not dividend-protected. Consistent with this deterrent to modifying distributions, Aboody and Kasznik (2008) find that firms that increased their dividends after JGTRRA modified their stock option and restricted stock compensation plans. In short, there were multiple reasons for firms to be hesitant about modifying their distribution policy following JGTRRA, even if individuals were seeking higher dividends than before the legislation.

III. RESEARCH DESIGN

System of Regression Equations

As discussed above, we jointly evaluate the changes in investor composition and distribution policy following the enactment of JGTRRA. We predict that individual investors rebalanced their portfolios so that a larger proportion of their shareholder income was in the form of dividends as opposed to share repurchases. We also expect that managers altered their distributions so that a larger proportion of their distributions were dividends and that this adjustment was increasing in the extent of individual ownership.

To jointly evaluate the impact of investor and manager responses to JGTRRA, we estimate a system of four equations (variables are defined below):

$$\begin{aligned} INSIDER = {} & \alpha_0 + \alpha_1 * POST + \alpha_2 * DIV\% + \alpha_3 * DIV\% * POST + \alpha_4 * S\&PRATING \\ & + \alpha_5 * AGE + \alpha_6 * SP500 + \alpha_7 * LIQUIDITY + \alpha_8 * BETA + \alpha_9 * IRISK \\ & + \alpha_{10} * MKTADJRET + \alpha_{11} * SALESGR + \alpha_{12} * R\&DINT; \end{aligned} \tag{1}$$

$$\begin{aligned} NONEXEC = {} & \delta_0 + \delta_1 * POST + \delta_2 * DIV\% + \delta_3 * DIV\% * POST + \delta_4 * S\&PRATING \\ & + \delta_5 * AGE + \delta_6 * SP500 + \delta_7 * LIQUIDITY + \delta_8 * BETA + \delta_9 * IRISK \\ & + \delta_{10} * MKTADJRET + \delta_{11} * SALESGR + \delta_{12} * R\&DINT; \end{aligned} \tag{2}$$

$$\begin{aligned} MF = {} & \gamma_0 + \gamma_1 * POST + \gamma_2 * DIV\% + \gamma_3 * DIV\% * POST + \gamma_4 * S\&PRATING + \gamma_5 * AGE \\ & + \gamma_6 * SP500 + \gamma_7 * LIQUIDITY + \gamma_8 * BETA + \gamma_9 * IRISK + \gamma_{10} * MKTADJRET \\ & + \gamma_{11} * SALESGR + \gamma_{12} * R\&DINT; \end{aligned} \tag{3}$$

$$\begin{aligned} DIV\% = {} & \beta_0 + \beta_1 * POST + \beta_2 * INSIDER + \beta_3 * INSIDER * POST + \beta_4 * NONEXEC \\ & + \beta_5 * NONEXEC * POST + \beta_6 * MF + \beta_7 * MF * POST + \beta_8 * RE + \beta_9 * PERM \\ & + \beta_{10} * TRANS + \beta_{11} * FCF + \beta_{12} * DYIELD + \beta_{13} * LEVERAGE + \beta_{14} * SIZE \\ & + \text{Industry Indicators}. \end{aligned} \tag{4}$$

To capture investor responses, the first three equations regress the percentage of the firm held by three individual investor groups on the percentage of the firm's payouts that are dividends (*DIV%*). There is one equation for each investor group: insiders (*INSIDER* in Equation (1)), other individuals (*NONEXEC* in Equation (2)), and mutual funds (*MF* in Equation (3)). The coefficients on the key variables of interest are α_3, δ_3, and γ_3, respectively. The fourth equation in the system flips the direction of the association and regresses the percentage of the firm's payouts that are dividends on each of the three investor groups. The fourth equation has three key variables of interest, one for each investor group in the period post-enactment: insiders (β_3), other individuals (β_5), and mutual funds (β_7).

Dividends-to-Payout Ratio Variable

The percentage of the firm's payouts that are dividends (*DIV%*) is an explanatory variable in the first three equations and the dependent variable in the final equation. *DIV%* is a ratio where the numerator is the sum of dividends over an eight-quarter period and the denominator is the sum of dividends and share repurchases over the same eight quarters. There are two eight-quarter periods for each firm. The first eight quarters are the eight quarters immediately before the fiscal quarter

in which JGTRRA was enacted.[11] The second eight quarters are the eight quarters immediately following the fiscal quarter in which the JGTRRA was enacted.[12] Consequently, each firm has two *DIV%* measures: one before enactment and one after enactment.

We aggregate distributions over a two-year period because, unlike regular, quarterly dividends, share repurchases are irregular. Therefore, focusing on the dividend-repurchase mix in a single quarter could introduce excessive noise. That said, in sensitivity tests reported later, we relax this aggregation requirement and report results on a quarterly basis. Inferences are largely unaltered.

We can measure repurchases in two ways. One option is total share repurchases. Another option is net repurchases—i.e., total share repurchases less stock issuances.[13] We use net repurchases because we are interested in the cash that the firm could have distributed as dividends. Fama and French (2001) note that dividends cannot substitute for repurchases in many situations. Firms need shares for executive compensation, stock option exercises, stock acquisitions, and funding employee stock ownership plans, among other things. Thus, consistent with Fama and French (2001), we measure net repurchases as the change in treasury stock.[14] If there is a net decrease in treasury stock, then we truncate our measure of repurchases at zero. For those firms that do not use the treasury stock method, we measure net repurchases as total repurchases from the statement of cash flows less decreases in preferred stock.

For post-enactment observations in the first three regressions, we interact *DIV%* with *POST*, a categorical variable that equals 1 for observations after the May 23, 2003 enactment (i.e., quarters after the second quarter of 2003). In these three tax clientele tests, positive coefficients on $DIV\% * POST$ (the coefficients are α_3, δ_3, and γ_3 in the system of equations) can be interpreted as evidence that, after passage of the JGTRRA, individuals rebalanced their portfolios by shifting toward stocks where dividends constituted a larger portion of total payouts. We also include *POST* as a separate variable in each regression equation to capture any other temporal change.

Investor Group Variables

We employ three variables to capture the portion of the firm owned by individual investors. The first individual ownership measure, *INSIDER*, is the percentage of shares held by directors and officers, as reported in Thomson Financial's Insider Filing Data.[15] Note that these shareholders play dual roles—as the managers setting distribution policy and as individual shareholders, often

---

[11] We exclude the enactment quarter (May 2003) because it is unclear which tax regime managers were contemplating when they issued dividends and repurchased shares during that quarter. We treat the first quarter of 2003 as a pre-enactment quarter, even though the legislation was retroactive to the beginning of the year. The reason is that passage of the legislation was uncertain until Vice-President Cheney's tiebreaking vote in the U.S. Senate in May. Sensitivity tests, detailed below, provide assurance that this classification is appropriate. In addition, inferences hold if we exclude any 2003 quarters from the pre-enactment period.

[12] Blouin and Krull (2009) show that share repurchases rose in 2005 as firms enjoyed a tax holiday for repatriating earnings from foreign subsidiaries. When we replicate our analysis excluding the 2005 quarters, inferences hold.

[13] For further detail, see the discussion in footnotes 5 and 6 of Boudoukh et al. (2007) and footnote 9 of Skinner (2008).

[14] Using treasury stock to measure repurchases is not without limitations. As Fama and French (2001) point out, using annual changes in treasury stock will fail to match a repurchase in one year and its reissuance in another year. This problem is mitigated in our research design because we combine two years of activity into one observation. However, even aggregation over two years cannot fully eliminate the potential mismeasurement.

[15] The reporting of holdings of insiders is mandated by Section 16 of the Securities Exchange Act of 1934, which applies to every person who is the beneficial owner of more than 10 percent of any class of equity security registered under Section 12 of the Exchange Act and each director and officer (collectively, "reporting persons" or "insiders") of the issuer of the security. On a different note, conclusions do not change if we limit *INSIDER* to direct holdings, excluding shares held by family members, trusts and corporations controlled by the insider, and similar related parties.

with large stockholdings and suffering from inadequate diversification.[16]

The second measure of individual ownership is *NONEXEC*, which is intended to measure all individual holdings, other than those by insiders or through mutual funds. Ideally, we would measure the number of shares for which dividends and capital gains are expected to flow through to individual tax returns—i.e., those shares held by individuals or flow-through entities (e.g., mutual funds, partnerships, trusts, S corporations, or limited liability corporations) whose income is reported on U.S. individual tax returns. This ideal measure would exclude those holdings for which the dividends and capital gains do not flow through to individual tax returns, such as tax-exempt organizations, corporations, foreigners, and tax-deferred accounts (e.g., qualified retirement plans, including pensions, 401(k), and IRAs). Unfortunately, the ideal measure does not exist. Thus, as in Ayers et al. (2002), Ayers et al. (2004), Blouin et al. (2003), and Dhaliwal et al. (2003), among others, we use 13-F filings to estimate the percentage of the firm held by individual shareholders. *NONEXEC* is 1 less (1) the percentage of shares that institutional investors own, as reported in 13-F filings and collected by Thomson Financial's Institutional Holdings database, (2) the percentage of shares held by non-officer/director beneficial owners as reported in Thomson Financial's Insider Filing data, and (3) *INSIDER*.

The third measure of individual ownership, *MF*, is the percentage of the firm owned by mutual funds, as reported in 13-F filings and collected by Thomson Financial's Institutional Holdings database. As mentioned above, this is an imperfect measure of individual ownership because mutual funds include both investments that are subject to personal taxes and investments that are not subject to personal taxes. Sometimes the dividends and capital gains realized by mutual funds are taxed at the individual level. At other times, distributions to mutual funds are exempt because the shares are held in deferred tax accounts, such as 401(k) or IRAs. We include *MF* as an attempt to capture all shareholder income that is taxed on personal tax returns. However, we recognize that the unobservable measurement error in *MF* (arising from the inclusion of non-individual owners) may lessen its usefulness.

Each of the three individual ownership measures serves as the dependent variable in one of the first three regressions. All three are explanatory variables in the fourth regression, which tests for firm responses. Positive coefficients on these variables are consistent with firms altering their payouts in response to changes in individual tax incentives. Specifically, a positive coefficient on $INSIDER * POST$ (β_3) in the fourth regression will be interpreted as evidence that, after passage of the JGTRRA, the percentage of payouts distributed as dividends was increasing in insider ownership. It seems likely that, when directors and officers hold large shares of a firm, payouts are likely to be particularly responsive to individual tax incentives. A positive coefficient on $NONEXEC * POST$ (β_5) in the fourth regression will be interpreted as evidence that, after passage, *DIV%* was increasing in the percentage of the firm held by non-executive shareholders. A positive coefficient on $MF * POST$ (β_7) in the fourth regression will be interpreted as evidence that, after passage, *DIV%* was increasing in the percentage of the firm held by mutual funds. The measurement error in *MF* (arising from the fact that non-individuals invest in mutual funds) should bias the coefficient on $MF * POST$ toward zero.

Control Variables

Theory is insufficiently rich to provide much guidance concerning the control variables in a system of equations where the dependent variables are investor composition and the mix of

[16] Brown et al. (2007) report that insiders were particularly influential among dividend initiators in 2003. They present evidence consistent with dividends crowding out repurchases in firms with large insider holdings. However, they find no such substitution or insider influence among companies that were paying dividends before JGTRRA, which is the group of firms that is the focus of our analysis.

dividends and repurchases. To our knowledge, no study to date models the non-tax variables that should vary with the dependent variables in this study. Thus, we control for a host of factors that have been found to be associated with the investor mix and the distribution mix.

For the three regression equations testing for clientele effects, we rely on Bushee (2001), who shows that the level of institutional ownership is associated with firm value, and Del Guercio (1996), who documents that institutional holders tend to hold investments that are more prudent. Hence, we include a number of control variables in the investor holdings regressions to capture firm value and the relative quality of the investment. Specifically, we include *SALESGR* as a proxy for firm growth, defined as the average sales growth over the three previous years. We include two proxies for firm risk: beta (*BETA*), which is included to control for systematic risk, and the standard deviation of the prior year's daily market model residuals (*IRISK*) to control for idiosyncratic risk. Market-adjusted returns over the prior year (*MKTADJRET*) is intended to control for firm performance, which has been found to be positively associated with institutional holdings. The S&P common stock rating (*S&PRATING*) and the number of years that the firm is covered by CRSP (*AGE*) are included to capture the relative quality of the underlying investment. The log of the prior year's average monthly volume divided by shares outstanding (*LIQUIDITY*) is included as a control for liquidity because institutional holders prefer more liquid securities. We also include whether a firm is listed on the S&P 500 (*SP500*) as a control because many index funds are required to hold these firms. Finally, we include R&D intensity (*R&DINT*), measured as research and development expenses divided by sales because Hessel and Norman (1992) report that some institutions are fixated on the R&D activity of the firm.

Concerning the fourth regression, where the dependent variable is *DIV%*, we include various measures that are known to affect either dividends or repurchases, although sensitivity tests show that results are largely robust to the set of control variables. First, we include lagged retained earnings scaled by lagged total assets (*RE*). A firm must have earnings and profits (as defined in the tax law) for its distributions to be taxed as dividends. Unfortunately, earnings and profits are unobservable, found only in confidential corporate tax returns. Thus, we use retained earnings as a proxy for earnings and profits. If firms with low or no retained earnings have fewer distributions that qualify as dividends, then *DIV%* should increase in *RE*. Consistent with this expectation and liquidity constraints, DeAngelo et al. (2006) report that firms with low or no retained earnings pay fewer dividends. Next, we include earnings in the model. Jagannathan et al. (2000) and Guay and Harford (2000) report that dividends are paid from permanent earnings, whereas repurchases are paid from transitory earnings. Dittmar and Dittmar (2004) contend that both are paid from permanent earnings, but agree that repurchases come from transitory earnings. Thus, we dichotomize earnings into a permanent component (*PERM*) and a transitory component (*TRANS*). We measure *PERM* with operating income and *TRANS* as the difference between net income and operating income, scaling both components by lagged assets.[17] Based on the conflicts in prior research, we make no prediction about the sign of *PERM*. However, we expect *DIV%* to decrease in *TRANS*. We also include a measure of the firm's payout capacity, free cash flow scaled by lagged assets (*FCF*). Dividends serve as a possible mechanism to reduce agency problems in firms with free cash flow (Jensen and Meckling 1976). Thus, we anticipate that *DIV%* is increasing in *FCF*. We include the lagged ratio of dividends to the market value of equity (*DYIELD*) to control for a firm's capacity to increase dividends. The higher the existing dividend yield, the less the firm will be able to increase its dividend in response to the passage of the JGTRRA. We expect *DIV%* to be increasing in *DYIELD*. We add lagged long-term debt, scaled by lagged assets (*LEVERAGE*), to

[17] Differences between permanent and transitory earnings include special items, other income, and discontinued operations.

control for cross-firm variation in capital structure. Finally, we include the natural logarithm of total assets (*SIZE*) to control for any size effects. We have no expectations about the sign of the *LEVERAGE* and *SIZE* coefficients.

IV. RESULTS

Sample Selection

We begin our tests for investor and manager responses following JGTRRA by drawing an initial sample from the 14,122 firms in the Compustat database between the second quarter of 2001 and the quarter preceding the one that includes May 2003. We then exclude (1) firms whose shares were not common or publicly traded, (2) firms that changed their fiscal year-end during our sample period, (3) financial institutions and insurance companies because regulatory constraints may inhibit management from altering the firm's payout policy, (4) firms with missing Compustat information, and (5) firms not in existence at any time from July 1, 2001, to June 30, 2005.

From the remaining 1,923 firms, we draw two balanced panels. Each firm must have one observation for the eight quarters preceding JGTRRA and one observation for the eight quarters following JGTRRA. Because *DIV%* is undefined if there is neither a dividend nor a repurchase, each firm must have at least one dividend or one repurchase, both before and after JGTRRA. The first sample (Dividend Payers Sample) is the 421 firms in the study that paid dividends sometime during the eight quarters immediately preceding JGTRRA and had either a dividend or repurchase (or both) during the eight quarters immediately following JGTRRA. The second sample (Dividend Payers and Repurchasers) is the 294 firms that both paid dividends and repurchased shares sometime during the eight quarters immediately preceding JGTRRA and had either a dividend or a repurchase (or both) during the eight quarters immediately following JGTRRA.

Our definition of *DIV%* forces us to exclude firms that distribute no profits to shareholders. We further limit our tests to firms that paid dividends at least once during the eight quarters preceding JGTRRA. The reasons for these limitations are twofold. First, firms that were paying dividends before passage paid 97 percent of the dividends issued in the four quarters following enactment. However, much of the JGTRRA research (e.g., Chetty and Saez 2005; Brown et al. 2007) has focused on the relatively narrow impact of JGTRRA on dividend initiation in 2003 alone.[18] Thus, we focus on dividend-paying firms as an under-studied, but more economically significant, portion of the economy. Second, dividend initiation and, to a lesser extent, a firm's first share repurchase convey more and different information to the markets than simply altering the amount of an ongoing stream of dividends or repurchases. Thus, firms that have a history of shareholder distributions can likely modify the dividend-repurchase mix at a lower cost than firms that have never paid dividends or repurchased shares. In fact, it is possible that some managers that had never paid dividends before JGTRRA considered initiating payouts in response to the changed tax incentives associated with JGTRRA, but decided that the costs of initiation exceeded the benefits of attracting individual investors who were now seeking more dividend income. By limiting our analysis to dividend-paying firms, we ensure that the potential costs of dividend initiation do not affect our estimates of managerial responsiveness to the changed individual tax incentives under JGTRRA. This both increases the power of our tests and removes an additional factor for which we would need to control if we included non-dividend-paying firms. One downside to limiting the sample to firms that were already paying dividends is that, if these firms were at their dividend capacity when JGTRRA was enacted, then they may have been unable to increase

[18] This is not to imply that initiations arising from the JGTRRA were unimportant. Initiators began issuing dividends that they might not have paid and likely will continue to issue dividends in the future because dividend payments are sticky and the market takes a dim view of dividend cuts.

their dividend payouts, even if they had wished to respond to the changed tax incentives for individual investors.

Another distinguishing factor about our analysis of JGTRRA is that we examine investor and firm responses through 2005. Desai and Dharmapala (2011) study portfolio rebalancing in 2003 only. Brown et al. (2007) and Aboody and Kasznik (2008) study changes in the dividend-repurchase mix for only two quarters after passage of the legislation. As Shevlin (2008) notes, examining such a short period requires quick response by firms and raises concerns about whether these studies miss important responses in 2004 and 2005. It seems unlikely that investors completely unwound their positions and firms had fully adjusted their notoriously sticky dividends and repurchase policy in response to the tax rate reductions during the seven months following enactment. By investigating a longer window, we permit a more deliberate response and can calibrate how long it took for the largest dividend tax rate reduction in history to fully permeate the economy.

Although we exclude non-dividend-paying firms from our primary tests, in the process of selecting the sample firms, we detect some initial evidence consistent with firms shifting from repurchases to dividends. Among the 1,923 firms from which we draw our samples, we find that 145 companies initiated dividends after JGTRRA, while only 30 firms omitted dividends (a net increase of 115 dividend issuers). Meanwhile, 222 firms began repurchasing after enactment while 370 companies stopped repurchasing (a net reduction of 148 repurchasers). We also find that 26 firms both initiated dividends and ceased repurchasing after passage of the JGTRRA, while only three firms omitted dividends and began repurchasing. Furthermore, among 702 firms that repurchased both before and after JGTRRA, 90 initiated dividends while only 10 firms omitted dividends, a net increase of 80 dividend issuers. On the other hand, among the 408 firms that paid dividends both before and after JGTRRA, 46 began repurchasing after passage, but 58 stopped buying back shares, a net decrease of 12 repurchasers. All of these comparisons are consistent with firms shifting from repurchases to dividends following enactment of the JGTRRA.

Descriptive Statistics

Table 1 provides means and medians for the regression variables, both before and after enactment of JGTRRA, for both samples and reports whether the pre- and post-enactment means and medians are significantly different.[19] The samples provide some evidence that dividends increased following JGTRRA (median total dividends and mean dividends per assets increased in both samples). However, there is less evidence that *DIV%* (i.e., dividends as a percentage of total shareholder distributions) increased after passage (only the difference in medians for the Dividend Payers and Repurchasers sample is significant). Recall, however, that we do not hypothesize about the overall impact of JGTRRA on dividends, repurchases, or *DIV%*. Rather, we predict a more positive association between *DIV%* and individual stock ownership following passage of JGTRRA.[20]

Both samples show decreases in non-executive individual holdings (*NONEXEC*) and in-

[19] Means are tested using a t-test of the means. The p-values for the medians are the larger p-values using the Wilcoxon and Kruskal-Willis tests.

[20] Special dividends are included in *DIV%*. Although the frequency of special dividends increased after the enactment of the JGTRRA, the relative magnitude of special dividends continued to be small as compared to regular dividends and repurchases. For our sample of 421 dividend payers and 294 dividend payers and repurchasers, special dividends constituted only 1 percent (0.4 percent) of aggregate payout made after (before) enactment of the JGTRRA.

TABLE 1

Descriptive Statistics
Means and Medians for Regression Variables

	Dividends 421 Firms (at least one dividend before enactment of the JGTRRA)				Dividend Payers and Repurchasers Sample 294 Firms (at least one dividend and one repurchase before enactment of the JGTRRA)			
	Means		Medians		Means		Medians	
	Before	After	Before	After	Before	After	Before	After
Total Dividends	152	200	17	27*	193	252	19	30*
Total Repurchases	140	251	2	3	200	327	13	9*
Dividends per Assets	0.03	0.04*	0.02	0.03	0.03	0.04*	0.02	0.03
Repurchases per Assets	0.03	0.04	0.00	0.00	0.05	0.05	0.01	0.01**
DIV%	0.78	0.78	0.89	0.89	0.68	0.72	0.75	0.81**
INSIDER	0.10	0.09	0.04	0.04	0.11	0.09	0.05	0.04
NONEXEC	0.35	0.30**	0.32	0.25**	0.35	0.30**	0.30	0.25**
MF	0.27	0.29*	0.26	0.29**	0.26	0.29*	0.26	0.29*
RE	0.32	0.32	0.32	0.32	0.34	0.33	0.34	0.36
PERM	0.16	0.18*	0.13	0.16*	0.18	0.20	0.16	0.18
TRANS	−0.07	−0.07	−0.06	−0.06	−0.08	−0.08	−0.07	−0.07
FCF	0.06	0.02**	0.08	0.03**	0.07	0.03**	0.08	0.04**
DYIELD	0.53	0.44	0.40	0.36	0.55	0.43	0.38	0.37
LEVERAGE	0.22	0.20*	0.22	0.18*	0.21	0.18	0.20	0.17
SIZE	6.85	7.03	6.74	6.93	6.94	7.12	6.74	6.94
S&PRATING	12	13**	16	16**	11	13**	16	16*
AGE	25	25	22	22	25	25	22	22
SP500	0.23	0.24	0	0	0.25	0.26	0	0
LIQUIDITY	−0.50	−0.21**	−0.36	−0.14**	−0.48	−0.23**	−0.30	−0.14**
BETA	0.63	0.92**	0.60	0.92**	0.65	0.91**	0.61	0.90**
IRISK	2.46	1.83**	2.29	1.67**	2.42	2.26**	1.81	1.63**
MKTADJRET	0.12	0.05**	0.12	0.03**	0.11	0.05**	0.11	0.03**

(continued on next page)

TABLE 1 (continued)

	Dividends 421 Firms (at least one dividend before enactment of the JGTRRA)				Dividend Payers and Repurchasers Sample 294 Firms (at least one dividend and one repurchase before enactment of the JGTRRA)			
	Means		Medians		Means		Medians	
	Before	After	Before	After	Before	After	Before	After
SALESGR	0.12	0.26**	0.06	0.19**	0.11	0.23**	0.06	0.18**
R&DINT	0.01	0.01	0	0	0.01	0.01	0	0

*, ** Indicates significance at the 5 percent and 1 percent levels, respectively, using a two-tailed test.

This table reports descriptive statistics for the 421 (294) firms that paid dividends (both dividends and repurchases) in the period surrounding the JGTRRA. Total Dividends is dividends as reported by Compustat (data16 × data61). If the firm uses the treasury stock method, then Total Repurchases is the change in treasury stock (data98). If there is a net decrease in treasury stock, then Total Repurchases is set to 0 (i.e., if there are net stock issuances). If the firm does not use the treasury stock method, then Total Repurchases is the repurchase amount from the statement of cash flows (data93) less decreases in preferred stock (change in data55 and data71). Dividends per Assets is Total Dividends scaled by lagged assets (data44). Repurchases per Assets is Total Repurchases scaled by lagged assets (data44).

Variable Definitions:

DIV% = Total Dividends over the sum of Total Dividends and Total Repurchases;
INSIDER = percentage of shares held by insiders as measured by holdings of directors and officers as reported in Thomson Financial's Insider Filing Data;
NONEXEC = 1 less the percentage of shares that are held by institutional investors as reported in 13-F filings and collected by Thomson Financial's Institutional Holdings database less *INSIDER* less non-officer/director beneficial owners as reported in Thomson Financial's Insider Filing data;
MF = percentage of the firm owned by mutual funds as reported in 13-F filings and collected by Thomson Financial's Institutional Holdings database;
RE = lagged retained earnings scaled by lagged assets (data58/data44);
PERM = operating income (data21−data5−data22) scaled by lagged assets (data44);
TRANS = net income scaled by lagged assets (data69/data44) less *PERM*;
FCF = income before extraordinary items (data8) plus interest expense (data22) less the change in the applicable balance sheet accounts: (assets (data44), liabilities (data54), debt (data45, data51)) scaled by lagged assets (data44);
DYIELD = lagged ratio of dividends per share to price, expressed in percentages (data16/data14);
LEVERAGE = lagged long-term debt (data51 + data45) scaled by lagged assets (data44);
SIZE = natural log of assets (data44);
S&PRATING = S&P Common Stock Ranking (SPCSR 7 = A+ etc.) ÷ 100;
AGE = number of years the firm is reported on CRSP ÷ 100 as of 2003;
SP500 = 1 if the firm is in the S&P500 Index, 0 otherwise;
LIQUIDITY = log of average monthly volume over shares outstanding for the prior year;

(continued on next page)

TABLE 1 (continued)

BETA = market beta estimated over the prior 12 months;
IRISK = unsystematic risk, which is estimated as the standard deviation of daily market model residuals over the prior year multiplied by 100;
MKTADJRET = market adjusted returns over the prior year expressed as a percentage;
SALESGR = average sales growth over the prior two years (data2/data44); and
R&DINT = R&D intensity estimated as R&D expense over lagged assets (data4/data44); missing R&D is set to 0.

creases in mutual funds holdings (*MF*).[21] Among the control variables, free cash flow (*FCF*), volume (*LIQUIDITY*), idiosyncratic risk (*IRISK*), and returns (*MKTADJRET*) fell after enactment for both samples. Stock ratings (*S&PRATING*), beta (*BETA*), and sales growth (*SALESGR*) increased after passage for both samples.

Preliminary Regression Results

To provide some perspective for the simultaneous equation results that we report shortly, Tables 2 and 3 show separate OLS regression results for each of the four equations. Table 2 shows selected coefficient estimates from the first three equations, where investor ownership percentages are the dependent variables and *DIV% * POST* is the variable of interest. Table 3 presents results

TABLE 2

OLS

Changes in Ownership by Insiders, Non-Executives, and Mutual Funds around JGTRRA Enactment

		Equation (1): *INSIDER* (Insider % Ownership)		Equation (2): *NONEXEC* (Non-Executive % Ownership)		Equation (3): *MF* (Mutual Fund % Ownership)	
	Pred.	**Dividend Payers**	**Dividend Payers and Repurchasers**	**Dividend Payers**	**Dividend Payers and Repurchasers**	**Dividend Payers**	**Dividend Payers and Repurchasers**
Intercept		−0.027	−0.040	0.273**	0.300**	0.414**	0.422**
POST		−0.013	−0.006	0.049	0.022	−0.020	−0.020
DIV%		−0.006	0.021	−0.008	−0.039	0.022	0.003
*DIV% * POST*	(+)	0.048	0.038	−0.004	0.003	−0.026	−0.011
S&PRATING		−0.029	−0.018	−0.144	−0.020	0.002*	0.184*
AGE		−0.084**	−0.063	0.064	0.072	−0.037	−0.076*
SP500		−0.001	−0.011	0.030	0.023	−0.093**	−0.082**
LIQUIDITY		−0.045**	−0.044**	−0.107**	−0.108**	0.094**	0.089**
BETA		−0.011	−0.003	−0.082**	−0.068**	0.030**	0.028**
IRISK		2.211**	2.543**	5.831**	2.968**	−3.785**	−2.783**
MKTADJRET		11.088*	15.281**	14.383	12.398	−7.804	−9.263
SALESGR		0.015	0.019	0.004	−0.019	−0.016	−0.004
R&DINT		0.133	0.191	−0.173	−0.423*	−0.291**	−0.255*
n		842	588	842	588	842	588

*, ** Indicate that the coefficients are significantly different from 0 at the 5 percent and 1 levels, respectively, using a one-tailed test where there are predictions; otherwise, using a two-tailed test.

This table reports estimated ordinary least squares regression statistics for the 421 (294) firms that paid dividends (both dividends and repurchases) in the period surrounding the JGTRRA. The dependent variables are the percentages of ownerships by insiders, non-executives, and mutual funds. Observations aggregate eight quarters before and after enactment of the JGTRRA. *POST* is a categorical variable that equals 1 if the dividend/repurchase was declared in the eight quarters after 5/23/03, 0 otherwise. All other variables are defined in Table 1.

[21] According to the Investment Company Institute (see http://www.ICI.org), from 2001 to 2005 mutual fund ownership of all publicly traded equity securities increased from 21 percent to 25 percent. In addition, over our sample period, individuals increased the proportion of their financial assets held by mutual funds from 40 percent in 2001 to 47 percent in 2005.

TABLE 3
OLS
Changes in Dividends and Repurchases around JGTRRA Enactment

		Equation (4): *DIV%*	
	Pred.	**Dividend Payers (n = 842)**	**Dividend Payers and Repurchasers (n = 588)**
Intercept		0.95**	0.94**
POST		−0.03	−0.03
INSIDER		−0.01	−0.02
*INSIDER * POST*	(+)	0.28*	0.31*
NONEXEC		0.02	−0.04
*NONEXEC * POST*	(+)	0.02	0.06
MF		0.03	−0.15
*MF * POST*	(+)	0.04	0.09
RE		0.11**	0.10**
PERM		−0.42**	−0.28**
TRANS		0.14	0.25
FCF		−0.16**	−0.16**
DYIELD		0.11	0.14
LEVERAGE		0.08	0.01
SIZE		−0.02**	−0.02*
Industry Indicators		Yes	Yes

*, ** Indicate that the coefficients are significantly different from 0 at the 5 percent and 1 percent levels, respectively, using a one-tailed test where there are predictions; otherwise, using a two-tailed test.

This table reports estimated ordinary least squares regression statistics for the 421 (294) firms that paid dividends (both dividends and repurchases) in the period surrounding the JGTRRA. The dependent variable is the percentage of shareholder payouts that are dividends. Observations aggregate eight quarters before and after enactment of the JGTRRA. All variables are defined in Tables 1 and 2.

for the fourth equation, where *DIV%* is the dependent variable and the ownership percentages of the three-investor groups are explanatory variables.

Starting with Table 2, we expect a positive coefficient on *DIV% * POST*, which will be interpreted as evidence that, following enactment, individual ownership increased for those firms that distributed larger portions of their profits as dividends. Using OLS, we find a positive coefficient on *DIV% * POST* (α_3) when *INSIDER* is the dependent variable. However, the coefficient is not significantly greater than 0 at the 5 percent level (although using a one-tailed test, the coefficient is significant at the 10 percent level for the Dividend Payers sample). Contrary to expectation, three of the four coefficients on *DIV% * POST* are negative when the dependent variable is *NONEXEC* (δ_3) or *MF* (γ_3), although none are significantly different from 0. In short, the OLS results in Table 2 provide no evidence that investors rebalanced their portfolios following enactment, shifting toward stocks that distributed a larger portion of their payout as dividends.

Among the control variables, two are significant in every regression. *LIQUIDITY* is negative when *INSIDER* or *NONEXEC* is the dependent variable and positive when *MF* is the dependent variable. The signs flip for *IRISK*, as its coefficient is always positive when *INSIDER* or *NONEXEC* is the dependent variable and always negative when *MF* is the dependent variable. These results are consistent with differences in the non-tax factors that matter to insiders and non-

executive individuals investing on their own account as compared to those that matter to mutual fund investors (see Bushee 2001).

Table 3 presents OLS summary statistics from regressing the ratio of dividends to total payouts (*DIV%*) on the measures of individual ownership and control variables. We predict positive coefficients on *INSIDER* $*$ *POST* (β_3), *NONEXEC* $*$ *POST* (β_5), and *MF* $*$ *POST* (β_7), consistent with firms altering their distribution policy to retain and attract individual investors, given the tax changes in JGTRRA.

We find that *INSIDER* $*$ *POST* coefficients are positive and significantly greater than 0 at the 5 percent level. We interpret these findings as evidence that firms with large holdings by directors and officers distributed a larger portion of their profits as dividends, after enactment, than they did before enactment. This is consistent with the individuals who set the dividend/repurchase policy (i.e., the directors and officers) modifying the distribution policy after enactment in a manner that is consistent with their own (and other individual) shareholders' interests. The *NONEXEC* $*$ *POST* and *MF* $*$ *POST* coefficients also are positive, but not significantly greater than 0 at conventional levels.[22] These initial results are consistent with managers responding to the altered tax incentives of directors and officers, but not other individual investors.

Among control variables, the coefficients on *RE* (retained earnings) are always significantly positive. The coefficients on *PERM* (operating income), *FCF* (free cash flow), and *SIZE* (total assets) are always significantly negative.

Primary Regression Results

We now discuss the results from estimating all four regression equations simultaneously. We find that coefficient estimates are larger and more significant using simultaneous estimation procedures, consistent with joint evaluation of investor and firm responses leading to superior estimates than were detected using separate estimations. We infer from the stronger results that investors and managers responded concurrently to the tax changes in JGTRRA.

Table 4 shows the results from estimating all four equations simultaneously using GMM. For brevity, we report only the coefficients for the key variables. Looking first at insider responses, we find that, when *INSIDER* is the dependent variable, the coefficient on *DIV%* $*$ *POST* (α_3) is significantly greater than 0 for the Dividend Payers sample at the 5 percent level. (It is significant at the 10 percent level using the Dividend Payers and Repurchasers sample.) This finding is consistent with directors and officers boosting their holdings in their own firms, after enactment, if their firms were distributing relatively large portions of their profits as dividends (i.e., it is consistent with a tax clientele effect).[23] The *DIV%* $*$ *POST* coefficient of 0.11 implies that an increase of one standard deviation in *DIV%* $*$ *POST*, or 0.28, would boost insider holdings by three percentage points. This is a substantial increase in holdings by directors and officers because they held only 10 percent of the firm, on average, before JGTRRA. In contrast, as in Table 2, we continue to find no evidence that non-executive individuals (δ_3) or mutual fund investors (γ_3) rebalanced their holdings by shifting toward firms distributing a higher portion of their profits as dividends, as compared with repurchases. Thus, we conclude that directors and officers were the only investor group changing their holdings in response to firm distribution policy.

---

[22] The *NONEXEC* $*$ *POST* coefficient is significant at the 10 percent level using a one-tailed test in the in the Dividend Payers and Repurchasers sample.

[23] Since the percentage of stock held by insiders was declining during the investigation period (see Table 1), insiders in high-dividend-paying firms may not have been buying far more shares in their own companies. Instead, perhaps these insiders simply sold few shares in their own companies during this period while insiders in lower dividend-paying firms were unloading large holdings.

TABLE 4

GMM Estimation of System of Equations
Changes in Ownership by Insiders, Non-Executives, and Mutual Funds and Changes in Dividends and Repurchases around JGTRRA Enactment

Equation	Dependent Variable	Explanatory Variable	Pred.	Dividend Payers (n = 842)	Dividend Payers and Repurchasers (n = 588)
(1)	*INSIDER*	*DIV% * POST*	(+)	0.11*	0.10
(2)	*NONEXEC*	*DIV% * POST*	(+)	0.09	−0.04
(3)	*MF*	*DIV% * POST*	(+)	−0.02	0.01
(4)	*DIV%*	*INSIDER * POST*	(+)	0.32**	0.42**
(4)	*DIV%*	*NONEXEC * POST*	(+)	0.19*	0.18*
(4)	*DIV%*	*MF * POST*	(+)	0.28*	0.33*

*, ** Indicate that the coefficients are significantly different from 0 at the 5 percent and 1 percent levels, respectively, using a one-tailed test.

This table reports estimated GMM estimates of the simultaneous equations' regression statistics for the 421 (294) firms that paid dividends (both dividends and repurchases) in the period surrounding the JGTRRA. The dependent variables are the percentages of ownerships by insiders, individuals, and mutual funds and the percentage of shareholder payouts that are dividends. Observations aggregate eight quarters before and after enactment of the JGTRRA. All variables are defined in Tables 1 and 2. Note that all control variables in Tables 2 and 3 are included in the estimation of the model but excluded from the table for brevity.

Looking at manager responses, we find strong evidence that firms held disproportionately by insiders, non-executive individuals, and mutual funds began to distribute a higher percentage of profits through dividends following enactment. When the dependent variable is *DIV%*, the coefficients on *INSIDER * POST* (β_3), *NONEXEC * POST* (β_5) and *MF * POST* (β_7) are all positive and significant at conventional levels for both samples. These results are consistent with managers modifying their payout policy in response to their individual owners' increased preference for dividends compared with repurchases, following passage of JGTRRA. In other words, as dividends became less tax-disadvantaged, firms appear to have provided more dividends as a percentage of their total payout. Brown et al. (2007) report similar findings but only for insiders with large holdings in firms that initiated dividends in 2003.[24] Our findings show that firm responses extended to a broader set of firms (non-initiating dividend payers) and individual investors.

Analyzing the economic significance of each coefficient using a one-standard deviation increase, we find that the *INSIDER * POST* coefficient of 0.32 (0.42) with the Dividend Payers (Dividend Payers and Repurchasers) sample implies a four (six) percentage point increase in *DIV%*—i.e., the percentage of payout distributed as dividends. With a pre-JGTRRA mean *DIV%* of 78 percent (68 percent), this suggests that, for every standard deviation increase in insider holdings, the ratio of dividend issuance to total distributions rose 5 percent (9 percent). Likewise, the *NONEXEC * POST* coefficient suggests a four percentage point increase in *DIV%* for both samples. Finally, the *MF * POST* coefficient of 0.28 (0.33) implies a four (five) percentage point increase in *DIV%* for the Dividend Payers (Dividend Payers and Repurchasers) sample. In short, all six coefficients across the two samples imply an economically significant change in distribution policy in response to altered investor preferences following JGTRRA.

[24] Brown et al. (2007) treat all 2003 dividend initiations as responses to JGTRRA, even though the legislation was highly controversial and passed by a single vote in the U.S. Senate in May.

We find that the GMM estimates are substantially stronger than those for OLS. This is consistent with our expectation that investor and manager responses are simultaneously determined. Evaluating only investors or only managers potentially explains some of the weak and conflicting results from prior JGTRRA studies (see reviews by Dharmapala [2009] and Shackelford [2009]). If we were to stop with the OLS results, then we would erroneously conclude that no investors rebalanced their portfolios and that the firm responses were limited to those companies with heavy insider ownership. We also would substantially understate the economic significance of legislation as estimated by using the regression coefficients.

Comparisons with Non-Event Periods

This section repeats the analyses detailed above using non-event periods. If the 2003 findings are related to JGTRRA, then we expect the coefficients from the non-event years to be different from the 2003 coefficients. We repeat the tests as if the actual May 2003 rate reductions had occurred in May 1994, and include the eight pseudo "pre-enactment" quarters prior to the quarter that includes May 1994 and the eight pseudo "post-enactment" quarters that follow the quarter that includes May 1994. We then repeat this seven more times, using May of each year as the pseudo event.[25]

Table 5 reports the GMM summary statistics for the key coefficients for each year for the Dividend Payers sample.[26] For example, estimating Equation (1) for 1994, where *INSIDER* is the dependent variable, the *DIV% * POST* coefficient is −0.04. For Equation (2), when *NONEXEC* is the dependent variable, the *DIV% * POST* coefficient is 0.23. For Equation (3), when *MF* is the dependent variable, the *DIV% * POST* coefficient is 0.01. For Equation (4), for the single firm response equation where the dependent variable is *DIV%*, the *INSIDER * POST* coefficient is 0.15, the *NONEXEC * POST* coefficient is 0.03, and the *MF * POST* coefficient is 0.02. Unlike the 2003 results for which four of the six coefficients are significantly positive, none of these six coefficients are significantly different from zero. However, at the same time, none of the 1994 coefficients are significantly different from their 2003 counterparts, suggesting that there may be nothing unique about 2003. Results for the remaining years are tabulated, and summary statistics for the coefficients from the non-event years are presented at the bottom of Table 5.

The findings confirm the inferences drawn from the primary tests in Table 4. The 2003 coefficient on *DIV% * POST*, when *INSIDER* is the dependent variable, exceeds its counterpart in every year except 2001. Using a t-test (sign test), we can reject at the 0.02 (0.07) level the null hypothesis that the 2003 coefficient would be randomly selected from a distribution formed by the eight coefficients from 1994 through 2001. Finding that the 2003 association was different provides further confirmation that insider holdings, following JGTRRA, were increasing in the dividend percentage of total payout. As further confirmation of the Table 4 results, we continue to find no indication of tax clientele movements by other individuals or mutual funds.

Turning to the manager responses to JGTRRA, we find that the 2003 *INSIDER * POST* coefficient always exceeds its counterparts in every year from 1994 to 2001 and significantly so from 1997 to 2000. The 2003 *NONEXEC * POST* coefficient is greater than its counterparts in every year except 1996, and is statistically significantly larger in two of the eight years. The 2003 *MF * POST* coefficient always exceeds its non-event counterparts, though not significantly. However, for all three 2003 coefficients, we can reject at the 1 percent level, using a t-test, that the 2003 coefficient would be randomly drawn from a distribution of the estimated coefficients from

---

[25] We stop with 2001 because any test of 2002 would result in the inclusion of some of the period following enactment of the JGTRRA, invalidating its use as a non-event, comparison period.

[26] Inferences are qualitatively unaltered if the Dividend Payers and Repurchasers sample is used.

TABLE 5

GMM Estimation of System of Equations
Changes in Ownership by Insiders, Non-Executives, and Mutual Funds and Changes in Dividends and Repurchases around Hypothetical, Non-Event Dates

	Equation (1)	Equation (2)	Equation (3)	Equation (4)	Equation (4)	Equation (4)
Dependent Variable	*INSIDER*	*NONEXEC*	*MF*	*DIV%*	*DIV%*	*DIV%*
Explanatory Variable	*DIV% * POST*	*DIV% * POST*	*DIV% * POST*	*INSIDER * POST*	*NONEXEC * POST*	*MF * POST*
2003 Estimated Regression Coefficients (from Table 4)	0.11*	0.09	−0.02	0.32**	0.19*	0.28*
Assume Law Had Passed in May of:						
1994	−0.04	0.23	0.01	0.15	0.03	0.02
1995	−0.02	0.12	−0.02	0.22*	0.08	0.02
1996	0.01	0.21	0.01	0.16	0.20*	0.24
1997	−0.02	0.04	0.02	**−0.07**	−0.02	−0.05
1998	0.09	−0.14	0.03	**−0.18**	**−0.20***	0.01
1999	−0.10	−0.12	0.04	**−0.22**	**−0.15**	0.13
2000	0.05	0.03	−0.02	**−0.17**	0.01	0.13
2001	0.15*	−0.01	−0.08	0.14	0.01	−0.01
Descriptive Stats for 1994−2001 Coefficients						
Mean	0.02	0.05	0.00	0.00	−0.01	0.06
Median	−0.01	0.04	0.01	0.04	0.01	0.02
Standard Deviation	0.08	0.14	0.04	0.18	0.13	0.10
Maximum	0.15	0.23	0.04	0.22	0.20	0.24
Prob. that 2003 coefficient could be randomly drawn from a distribution of 1994−2001 coefficients using t-test (sign test)	0.02 (0.07)	0.38 (0.73)	0.21 (0.38)	0.01 (0.01)	0.01 (0.07)	0.01 (0.01)

*, ** Indicate that the estimated regression coefficient is significantly different from 0 at the 5 percent and 1 percent level, respectively. Bold coefficients are significantly different from their 2003 counterparts at the 5 percent level.

(continued on next page)

TABLE 5 (continued)

This table reports estimated GMM estimates of the simultaneous equations' regression statistics for the 421 firms that paid dividends in the period surrounding the JGTRRA from Table 4. The dependent variables are the percentages of ownerships by insiders, individuals, and mutual funds and the percentage of shareholder payouts that are dividends. The investigation periods are quarters surrounding various non-event periods. Sample firms issue at least one dividend during the two years preceding the assumed date of legislative passage. *POST* is a categorical variable that equals 1 if the dividend/repurchase was declared in the eight quarters after May 23 of each year, 0 otherwise. All variables are defined in Tables 1 and 2. Note that all control variables in Tables 2 and 3 are included in the estimation of the model but excluded from the table for brevity.

the non-event years. We can similarly reject at the 1 percent (7 percent) level, using the sign test, for the *INSIDER* ∗ *POST* and *MF* ∗ *POST* (*NONEXEC* ∗ *POST*) coefficients. Together, these results support our earlier conclusions that managers of firms with disproportionately large individual investor holdings modified their distribution policy more than did other firms. These sensitivity tests provide comfort about our earlier inferences, providing evidence that the associations in 2003 between distribution policy and individual ownership measures differ from the relations between those variables in the prior decade.

There is a caveat for this robustness check. The capital gains tax rate fell from 28 percent to 20 percent in 1997 without any change in the dividend tax rate. Using the same logic developed in this study, we would have predicted a shift from dividends to repurchases following that rate reduction, albeit of a lesser extent because the rate change was more modest. If that change in the distribution mix did occur, then this sensitivity test using non-event periods would bias in favor of our finding that the 2003 coefficients were more positive than those in previous years. The reason is that the tax incentives facing individual investors after the rate reductions in 1997 would have called for the mix of distributions to shift toward repurchases and away from dividends. In other words, 1997 was not really a non-event year.

Therefore, if firms responded accordingly, then we would expect the coefficients on the individual tax measures for 1997 and perhaps 1998 to be negative. Consistent with firms modifying their payout policy if individuals held disproportionately large interests, the coefficients on *INSIDER* ∗ *POST*, *NONEXEC* ∗ *POST*, and *MF* ∗ *POST* (when *DIV%* is the dependent variable) turn negative for the first time in 1997. Thus, firm responses to changing tax incentives following the 1997 Act may have biased in favor of our rejecting the null hypothesis that the distribution mix did not change after 2003. However, when we exclude the 1997 coefficients in our comparison of the 2003 coefficients to those in the other non-event periods, we still find that the 2003 coefficient estimate remains significantly different from those in the other years. Furthermore, if managers did indeed respond to the 1997 tax cuts by increasing the repurchase portion of their distributions, then these results provide further evidence that managers consider the personal taxes of their shareholders in issuing dividends and repurchasing shares.

How Quickly Did Investors and Managers Respond to JGTRRA?

All of the tests reported to this point aggregate dividends and repurchases over the eight quarters before enactment and the eight quarters after enactment, creating two observations (before and after JGTRRA) for each firm. As discussed above, the reason for aggregation is that, unlike regular, quarterly dividends, repurchases are irregular events, which can lead to highly volatile quarterly measures of *DIV%*. In this section, we relax this restriction and treat each quarter as a different observation, resulting in 16 observations for each firm. The reason we shift to quarterly measures here is to enable us to pinpoint the time when investors and managers responded to JGTRRA.

To get GMM quarterly coefficient estimates for the same key six variables examined throughout this study, we suppress the intercept and include a categorical variable for each quarter from the earliest quarter (eight quarters before enactment—the quarter including May 2001) to the most recent quarter (eight quarters after enactment—the quarter including May 2005). We then interact the categorical variable for each of the 16 quarters with the six variables of interest.

Table 6 reports 16 quarters of coefficient estimates for the six key variables for the Dividend Payers sample. We find little change across the quarters for the three tax clientele regressions. That is not surprising for estimates when *NONEXEC* or *MF* is the dependent variable because we find no evidence anywhere in this study that suggests that non-executive individuals or mutual fund investors rebalanced their portfolios in favor of stocks that pay a larger portion of their payouts through dividends. However, because we find some support for clientele effects among insiders, it

TABLE 6

GMM Estimation of System of Equations
Quarterly Changes in Ownership by Insiders, Non-Executives, and Mutual Funds and Changes in Dividends and Repurchases around JGTRRA Enactment

	Equation (1)	Equation (2)	Equation (3)	Equation (4)	Equation (4)	Equation (4)
Dependent Variable	*INSIDER*	*NONEXEC*	*MF*	*DIV%*	*DIV%*	*DIV%*
Explanatory Variable	*DIV% * EVENT*	*DIV% * EVENT*	*DIV% * EVENT*	*INSIDER * EVENT*	*NONEXEC * EVENT*	*MF * EVENT*
−8	−0.01	−0.02	0.02	−0.17	−0.18	−0.08
−7	0.01	−0.01	0.03	−0.07	−0.08	0.20
−6	−0.03	−0.02	0.02	−0.32**	−0.25*	−0.15
−5	−0.01	0.01	0.01	−0.18	−0.13	−0.09
−4	0.02	−0.07**	0.03*	0.08	−0.15	−0.01
−3	−0.01	−0.05*	0.03	−0.01	−0.05	−0.01
−2	0.01	−0.05	0.04*	−0.06	−0.14	0.01
−1	−0.00	−0.01	−0.00	−0.09	−0.14	−0.38*
Enactment Quarter	NA	NA	NA	NA	NA	NA
+1	−0.00	−0.01	0.02	−0.04	−0.03	0.06
+2	0.02	0.00	0.01	0.27*	0.06	0.18
+3	0.02	−0.02	0.02	0.30*	0.10	0.23
+4	0.01	−0.01	0.01	0.19	0.07	0.22
+5	−0.02	0.02	0.02	0.20	0.34**	0.52**
+6	−0.01	0.00	−0.02	−0.03	−0.11	−0.11
+7	0.04*	−0.03	−0.01	−0.25*	−0.11	0.00
+8	0.01	0.01	−0.01	−0.26*	0.19	0.14

*, ** Indicate that the coefficients are significantly different from 0 at the 5 percent and 1 percent level, respectively, using two-tailed tests.

This table reports estimated GMM estimates of the simultaneous equations' regression statistics for the 421 firms that paid dividends in the period surrounding the JGTRRA. The dependent variables are the percentages of ownerships by insiders, individuals, and mutual funds and the percentage of shareholder payouts that are dividends. Each of the 16 quarters surrounding the enactment of the JGTRRA is treated as a separate observation. The 421 sample firms issued dividends at least once during the two years before JGTRRA. *EVENT* is a categorical variable that equals 1 if the dividend was declared in applicable quarter, 0 otherwise. All variables are defined in Tables 1 and 2. Note that all control variables in Tables 2 and 3 are included in the estimation of the model but excluded from the table for brevity.

is a bit surprising that we cannot detect any cross-quarter changes in coefficient estimates.

Conversely, for manager responses to JGTRRA, we find a sharp increase in the quarterly coefficient estimates during the second quarter following passage, which would be the quarter ending December 31, 2003, for most companies (i.e., those with March, June, September, and December year-ends). In eight of the nine preceding quarters, including the quarter immediately following the quarter of enactment (quarter +1), the *INSIDER* coefficient is negative. Beginning with quarter +2, the *INSIDER* coefficient is positive in all but one quarter, and is significantly positive in four of the seven quarters. Likewise, the quarterly pattern for the *NONEXEC* coefficients shows a similar break in quarter +2. The *NONEXEC* coefficient is negative for all nine quarters before quarter +2. The coefficient turns positive in that quarter and remains positive in four of the remaining six quarters. Similarly, the *MF* coefficient is negative in six of the eight pre-enactment quarters and negative only once thereafter. The *MF* coefficient is actually positive in quarter +1, but its coefficient triples from quarter +1 to quarter +2.

We conclude from the quarterly findings that companies with disproportionately large individual ownership began to substantially adjust their payout policy in response to JGTRRA during the last quarter of 2003 and the changes were sustained for the remainder of the investigation period. A delay of one quarter before firms responded to the tax incentives of individual investors is consistent with firms being unable or unwilling to respond to the new tax rates in the quarter immediately following passage. Perhaps they had already made their dividend and repurchase decisions for the third quarter of 2003 by May 2003. However, by the last quarter of 2003, it appears that firms held disproportionately by individuals were beginning to shift at the margin from repurchases to dividends.

This delay of one quarter may partially explain Brown et al.'s (2007) inability to find non-dividend initiators substituting dividends for share repurchases in 2003 and Aboody and Kasznik's (2008) failure to link repurchases, stock options, and individual ownership. Both studies treat all dividends paid and shares repurchased in 2003 as post-enactment payouts.[27] If few firms were adjusting their distribution policy before the last quarter of 2003 and, in fact, firms were making distribution choices before then, as though prior tax law applied (as implied by the negative coefficients), then it is understandable that their tests would have struggled to detect any movement from repurchases to dividends.

V. CLOSING REMARKS

This study extends our understanding of the effects of shareholder taxes on firm payout policy by estimating a system of equations to quantify the investor and managerial responses to the unprecedented 2003 cuts in dividend and capital gains tax rates. We hypothesize that, in response to the legislation, individual investors (the only ones affected by the rate reductions) rebalanced their portfolios to increase their dividend income, while firms boosted the portion of their profits that they returned as dividends. Comparing firm-level individual ownership and dividend-repurchase mix, before and after 2003, we find evidence consistent with directors and officers

[27] Our quarterly results, indicating no response to JGTRRA before the second half of 2003 and very little until the last quarter of 2003, raise concerns about the pre/post JGTRRA classifications in Brown et al. (2007), Chetty and Saez (2006, 2005), and Aboody and Kasznik (2008). Those studies treat all dividends paid and shares repurchased in 2003 as responses to JGTRRA, even those declared months before the May 23rd passage of the legislation, (e.g., Microsoft's initial dividend announcement on January 7, 2003). By using the day that the dividends were paid, they even include some dividends declared in 2002, well before President Bush ever mentioned possible dividend tax relief in January 2003. Brav et al. (2008) state that it is implausible that firms were so clairvoyant that they declared tax-motivated dividends months before passage. Among other factors, President Bush's initial comments were vague and preliminary, and weeks passed before details of his proposal emerged. Furthermore, passage of the highly controversial legislation was uncertain until Vice-President Cheney cast a tie-breaking vote in the U.S. Senate to gain passage of the legislation.

rebalancing their portfolios, but not other individual investors. We also find evidence consistent with those firms with large individual ownership boosting the dividend portion of their total payouts. The payout adjustments began a few months following enactment of the tax cuts. The regression coefficient estimates imply that both the portfolio rebalancing and distribution policy changes following JGTRRA are economically significant.

To our knowledge, this is the first study to evaluate investor and managerial responses to JGTRRA using simultaneous equations. We find that the results are much stronger than they would have been had we used standard OLS estimates of separate regressions, consistent with investors and managers acting concurrently. Both statistical and economic significance increased substantially using GMM. Furthermore, this is the most comprehensive analysis of the JGTRRA to date. Other studies focus on either investor responses or managerial responses; we study both. We analyze the set of firms that pay almost all dividends (i.e., non-dividend initiators), three different groups of individual investors (insiders, non-executive individuals, and mutual funds), and the 16 quarters centered on passage of the legislation. The methodological and sample enhancements enable us to construct a more powerful test of the association between shareholder taxes and corporate distribution policy.

Finally, our results suggest that additional research is needed to understand the role of insiders in the interaction of shareholder taxes and distribution policies. We find that insiders are the only individuals who rebalanced their portfolio in response to the rate changes. We also find that firms were particularly responsive to the changed tax incentives if directors and officers held large positions. These results complement Brown et al.'s (2007) finding that insiders played key roles in initiating dividends at the expense of share repurchases in 2003, which suggests that corporate governance needs to be introduced into analyses of the shareholder tax-payout choice. Questions that future research could explore include the following: To what extent are insiders motivated by their own personal tax considerations when setting distribution policy? How do the tax incentives of insiders affect the returns to other investors? From a distribution policy perspective, did other individual investors benefit from holding stocks in companies with high insider interests while non-individual shareholdings suffered? To what extent does compensation of executives with stock affect a firm's distribution policy? To what extent do changes in shareholder taxes affect a firm's value differently, depending on the extent of its insider holdings? We look forward to answers to these and similar questions in future studies.

REFERENCES

Aboody, D., and R. Kasznik. 2008. Executive stock-based compensation and firms' cash payout: The role of shareholders' tax-related payout preferences. *Review of Accounting Studies* 13 (2–3): 216–251.

Auerbach, A. J., and K. A. Hassett. 2007. The 2003 dividend tax cuts and the value of the firm: An event study. In *Taxing Corporate Income in the 21st Century*, edited by A. J. Auerbach, J. R. Hines, Jr., and J. Slemrod, 93–126. London, U.K.: Cambridge University Press.

Ayers, B. C., B. Cloyd, and J. Robinson. 2002. The effect of shareholder-level dividend taxes on stock prices: Evidence from the Revenue Reconciliation Act of 1993. *The Accounting Review* 77 (4): 933–947.

Ayers, B., C. Lefanowicz, and J. Robinson. 2004. The effect of shareholder-level capital gains taxes on the structure of corporate acquisitions. *The Accounting Review* 79 (4): 859–887.

Bagwell, L. S., and J. B. Shoven. 1989. Cash distributions to shareholders. *The Journal of Economic Perspectives* 3 (Summer): 129–149.

Berhheim, B. D. 1991. Tax policy and the dividend puzzle. *The RAND Journal of Economics* 22 (4): 455–476.

Blouin, J. 2009. Discussion of: Dividend tax clienteles: Evidence from tax law changes. *The Journal of the American Taxation Association* 31 (1): 23–28.

———, and L. Krull. 2009. Bringing it home: A study of the incentives surrounding the repatriation of foreign earnings under the American Jobs Creation Act of 2004. *Journal of Accounting Research* 47 (4): 1027–1059.

———, J. Raedy, and D. Shackelford. 2003. Capital gains taxes and equity trading: Empirical evidence. *Journal of Accounting Research* 41 (4): 611–651.

Boudoukh, J., R. Michaely, M. Richardson, and M. Roberts. 2007. On the importance of measuring payout yield: Implications for empirical asset pricing. *The Journal of Finance* 62 (2): 877–915.

Brav, A., J. Graham, C. Harvey, and R. Michaely. 2008. Managerial response to the May 2003 dividend tax cut. *Financial Management* 37 (4): 611–624.

Brown, J., N. Liang, and S. Weisbenner. 2007. Executive financial incentives and payout policy: Firm responses to the 2003 dividend tax cut. *The Journal of Finance* 62 (4): 1935–1965.

Bushee, B. J. 2001. Do institutional investors prefer near-term earnings over long-run value? *Contemporary Accounting Research* 18 (2): 207–246.

Chetty, R., and E. Saez. 2010. Dividend and corporate taxation in an agency model of the firm. *American Economic Journal: Economic Policy* 2 (3): 1–31.

———, and ———. 2005. Dividend taxes and corporate behavior: Evidence from the 2003 dividend tax cut. *The Quarterly Journal of Economics* 120 (3): 791–833.

———, and ———. 2006. The effects of the 2003 dividend tax cut on corporate behavior: Interpreting the evidence. *American Economic Review Papers and Proceedings* 96 (2): 124–129.

DeAngelo, H., L. DeAngelo, and R. Stulz. 2006. Dividend policy and the earned/contributed capital mix: A test of the lifecycle theory. *Journal of Financial Economics* 81 (2): 227–254.

Del Guercio, D. 1996. The distorting effect of the prudent-man laws on institutional equity investments. *Journal of Financial Economics* 40 (1): 31–62.

Desai, M., and D. Dharmapala. 2011. Dividend taxes and international portfolio choice. *The Review of Economics and Statistics* 93 (1): 266–284.

Dhaliwal, D., M. Erickson, and R. Trezevant. 1999. The effect of dividend initiations on institutional holdings: A test of the tax-clientele hypothesis. *National Tax Journal* 52 (2): 179–194.

———, and O. H. Li. 2006. Investor tax heterogeneity and ex-dividend day trading volume. *The Journal of Finance* 61 (1): 463–490.

———, ———, and R. Trezevant. 2003. Is a dividend tax penalty incorporated into the return on a firm's common stock? *Journal of Accounting and Economics* 35 (2): 155–178.

Dharmapala, D. 2009. The impact of taxes on dividends and corporate financial policy: Lessons from the 2000s. In *Tax Policy: Lessons from the 2000s*, edited by A. D. Viard, 199–233. Washington, D.C.: The AEI Press.

Dittmar, A. K., and R. F. Dittmar. 2004. Stock repurchase waves: An explanation of the trends in aggregate corporate payout policy. Working paper, University of Michigan.

Fama, E., and K. French. 2001. Disappearing dividends: Changing firm characteristics or lower propensity to pay. *Journal of Financial Economics* 60 (1): 3–43.

Feldstein, M., and J. Green. 1983. Why do companies pay dividends? *The American Economic Review* 73 (1): 17–30.

Gordon, R., and M. Dietz. 2008. Dividends and taxes. In *Institutional Foundations of Public Finance: Economic and Legal Perspectives*, edited by A. Auerbach, and D. Shaviro, 204–224. Boston, MA: Harvard University Press.

Grullon, G., and R. Michaely. 2002. Dividends, share repurchases and the substitution hypothesis. *The Journal of Finance* 57 (4): 1649–1684.

Guay, W., and J. Harford. 2000. The cash-flow permanence and information content of dividend increases versus repurchases. *Journal of Financial Economics* 57 (3): 385–415.

Guenther, D., and R. Sansing. 2010. The effect of tax-exempt investors and risk on stock ownership and expected returns. *The Accounting Review* 85 (3): 849–875.

Hessel, C., and M. Norman. 1992. Financial characteristics of neglected and institutionally held stocks. *Journal of Accounting, Auditing & Finance* 7 (3): 313–330.

Jagannathan, M., C. P. Stephens, and M. S. Weisbach. 2000. Financial flexibility and the choice between dividends and stock repurchases. *Journal of Financial Economics* 57 (3): 355–384.

Jensen, M., and W. Meckling. 1976. Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics* 3 (4): 305–360.

Miller, M., and F. Modigliani. 1961. Dividend policy, growth and the valuation of shares. *The Journal of Business* 34 (4): 411–433.

Moser, W., and A. Puckett. 2009. Dividend tax clienteles: Evidence from tax law changes. *The Journal of the American Taxation Association* 31 (1): 1–22.

Shackelford, D. 2009. A response to Dhammika Dharmapala. In *Tax Policy: Lessons from the 2000s*, edited by A. D. Viard, 234–247. Washington, D.C.: The AEI Press.

Shevlin, T. 2008. Discussion of: Executive stock-based compensation and firm's cash payout: The role of shareholders' tax-related payout preferences. *Review of Accounting Studies* 13 (2–3): 252–265.

Skinner, D. J. 2008. The evolving relation between earnings, dividends and stock repurchases. *Journal of Financial Economics* 87 (3): 582–609.

THE ACCOUNTING REVIEW
Vol. 86, No. 3
2011
pp. 915–943

American Accounting Association
DOI: 10.2308/accr.00000040

How a Systems Perspective Improves Knowledge Acquisition and Performance in Analytical Procedures

Billy E. Brewster
The University of Texas at Arlington

ABSTRACT: Auditors are required to understand dynamic business environments as part of the performance of analytical procedures. Prior evidence, though, suggests that auditors have difficulty understanding such environments. This study reports an experimental investigation of techniques that help auditors to identify incorrect management representations by developing expectations that are both accurate and adaptive to changing business conditions. My predictions suggest that analyzing a dynamic client environment through a systems perspective enhances information-processing ability, which then improves both evidence discrimination and the assimilation of new audit evidence. Results reveal that participants taking a holistic systems-based view of a client environment develop more coherently organized mental models that increase their likelihood of identifying management representations that are inconsistent with industry evidence. Furthermore, these participants more efficiently use their information-processing ability, thereby improving assimilation of newly learned evidence to understand how changing business conditions affect their initial expectations.

Keywords: *analytical procedures; knowledge organization; learning; mental models.*

Data Availability: *Contact the author for data availability.*

I. INTRODUCTION

When performing analytical procedures, auditors should treat any discrepancies between pre-developed expectations and management representations as indicators of heightened misstatement risk (Bell et al. 2005). Auditors develop expectations by understanding how a client's economic environment (entity-level evidence) relates to that client's financial statements (AICPA 2006; Knechel 2007). Because auditors compare their expectations about how financial and nonfinancial measures change over time to management representations of the same,

This paper is based on my dissertation at The University of Illinois at Urbana–Champaign. I thank my committee members, Mark Peecher (Chair), Paul Beck, Bruce Hannon, Joel Pike, and Ira Solomon for their insightful guidance on this project. I also thank Brian Ballou, Dain Donelson, Annie Farrell, Jon Grenier, Pat Hopkins, Steve Kachelmeier, Lisa Koonce, Susan Krische, Justin Leiby, Laureen Maines, Jeff Miller, Lillian Mills, Drew Reffett, Kenny Reynolds, Sandra Vera-Muñoz, and workshop participants at The University of Illinois at Urbana–Champaign, Indiana University, Miami University, University of Notre Dame, The University of Texas at Arlington, and The University of Texas at Austin.

Editor's note: Accepted by Kathryn Kadous.

Submitted: October 2008
Accepted: October 2010
Published Online: April 2011

the complex and evolving nature of today's business environment increases auditors' needs to develop expectations that are not only accurate, but also adaptable to changing business conditions (Bell et al. 1997). The purpose of this study is to investigate techniques for examining entity-level evidence that improve analytical procedures performance in both corroborating management representations and accurately updating expectations when business conditions change.

While current audit approaches and standards require auditors to understand complex business environments, laboratory results suggest that even experienced decision-makers have difficulty learning in, and making optimal decisions about, such settings (Sterman 1989; Diehl and Sterman 1995). Consistently, auditors, particularly inexperienced auditors, have difficulty linking entity-level evidence to management representations (Curtis and Turley 2007), and this inability is often a significant component of audit failure (SEC 2003; Peecher et al. 2007). At least a part of this difficulty is likely due to human cognitive limitations inhibiting the auditor's ability to build accurate and sufficiently detailed mental models that take into account how economic events flow down to affect the client's financial statements (Bell et al. 2002, 2005).[1] Of particular difficulty is understanding the nonlinear movement of entity-level evidence. Since auditors develop expectations by cognitively simulating the time-series behavior of how entity-level evidence relates to the financial statements, these cognitive limitations hinder auditors' ability to develop accurate expectations that can be used as a benchmark for evaluating management representations.

Auditing and management theorists suggest one way to improve performance is to train auditors to adopt a systems-based perspective when analyzing a client's economic environment (Bell et al. 1997; Sterman 2000; Peecher et al. 2007). This systems view involves developing, and organizing, a mental model of the client environment by holistically evaluating causal relationships and how they change over time. Drawing on psychology theory, I predict that organizing a mental model using a systems view will increase information-processing ability, thereby improving the accessibility of auditor's pre-existing knowledge of entity-level evidence. This enhanced processing should lead to more accurate and adaptable nonlinear expectations. Without a systems view, cognitive limitations are expected to force auditors into a reductionist view that considers environment components in isolation and simulates more predictable, and often incorrect, linear expectations.

Prior accounting research provides evidence that using systems-based concepts such as holistic templates (Choy and King 2005) and understanding circular causality (Hecht 2010) can improve decisions. However, these studies do not place the decision-maker in a complex setting that strains cognitive-processing resources and requires using pre-existing knowledge to simulate outcomes for evaluating financial information. Such an examination is needed since practitioners note that understanding and using knowledge from complex business environments is an essential step in solving many ambiguous auditing problems (PwC 2003). Also, no study shows how auditors learn, build, and use systems-consistent mental models of their clients or the cognitive factors and processes that moderate performance improvements.

To test this prediction, I conduct a laboratory experiment in which novices perform an analytical procedures task, requiring them to learn about a dynamically complex client environment. The combination of novices and a complex environment allows me to test the theories underlying this study by placing participants in a setting that would overwhelm their cognitive-processing ability unless they have a well-organized mental model of the client. I prompt participants to analyze the client using either a systems-based perspective, to enable a more coherent organization

[1] Mental models are holistic cognitive representations of a specific phenomenon and are constructed based on an individual's understanding of how pieces of information link and interrelate (Rickheit and Sichelschmidt 1999; Bell et al. 2002). Mental models are comprised of one or many knowledge structures, and the model's level of detail and accuracy is subject to both the individual's understanding of the phenomena and cognitive-processing constraints.

of entity-level evidence, or a reductionist-based perspective, leaving the participants to organize such evidence intuitively. I then manipulate whether management provides an explanation for an account fluctuation that is either relatively more consistent or inconsistent with the entity-level evidence. Finally, to learn how auditors update their mental models for changing conditions, I introduce new entity-level evidence that should change the nature of the auditors' existing expectations.

Consistent with predictions, I find that systems participants provide more coherently organized mental simulations of the environment's causal relations. They perform these improved simulations without increasing their cognitive effort, suggesting that the systems view helps participants overcome the previously mentioned cognitive limitations by enhancing their information-processing ability (Clark et al. 2006). Systems participants are more likely than reductionist participants to incorporate key nonlinear industry attributes into their expectations, and are also more likely to identify management representations that are inconsistent with entity-level evidence. Finally, the systems participants who have coherently organized mental models are better able to more accurately update their expectations when presented with new entity-level evidence.

This study extends the existing literature on improving knowledge acquisition and performance during analytical procedures. Prior studies show that auditors' knowledge acquisition for analytical procedures is improved through feedback or increasing cognitive effort (Bonner and Walker 1994; Earley 2001; Moreno et al. 2007). However, in an audit setting, feedback is not always readily available and increasing cognitive effort is not always effective due to the complexity of some analytical procedures tasks. Consequently, audit firms and educators have explored methods to improve auditor mental models, but these methods expect auditors to make the essential causal linkages and simulations of the client environment that prior laboratory evidence suggests is difficult to do without assistance (Bell et al. 1997; Bell et al. 2002). Indeed, previous studies show that such approaches can actually introduce biases into the judgment process (O'Donnell and Shultz 2005). I demonstrate that taking a systems view improves participants' analytical procedures performance both initially (i.e., when gathering entity-level evidence and analyzing management representations), and when presented with new, incoming evidence.

Furthermore, I provide insight into how auditors build mental models by examining the factors that moderate the relationship between the organization (structure) of auditors' entity-level knowledge and their information-processing ability. This relationship accounts for substantial variation in task performance. While Borthick et al. (2006) find that better-developed knowledge structures improve performance, there are few studies examining how auditors build these structures and use them to form judgments. In addition, theoretical learning frameworks suggest that knowledge structure and processing ability are related (Kyllonen and Woltz 1989; Libby 1995), but to date, there is little empirical evidence of the true nature of this association (Ceci et al. 2003). I find that coherently organized knowledge structures enhance information-processing ability, which then improves the likelihood of accurately updating existing knowledge structures with new entity-level evidence. Understanding such a relationship allows audit firms and educators to tailor instruction and training to exploit such learning improvements.

In Section II, I describe relevant research and develop hypotheses. Section III presents the experimental method used to test the hypotheses, and Section IV discusses the results of the experiment. Section V concludes.

II. THEORY AND HYPOTHESES

Background on Learning and Knowledge Acquisition

Understanding how auditors learn and update expectations using a systems view requires examining the cognitive factors that facilitate learning entity-level evidence. Learning is the acquisition and organization of *knowledge* through *working memory* (Bonner 2007). *Working*

memory, the cognitive system for holding and manipulating information, is the intersection between long-term memory and information received through the senses, and is a significant component of intellectual and problem-solving ability (Baddeley 1998; Conway et al. 2003). *Knowledge* is information that is stored and retrieved from long-term memory (Libby 1995; Bonner 2007), and is comprised of content (specific information pieces) and the organization, or structure, of that content. When forming judgments, individuals retrieve knowledge from long-term memory and manipulate it within working memory.

Humans have difficulty learning and performing complex tasks due to limitations within working memory. Individuals can cognitively process only a few information items at a time, and their ability to hold and manipulate items in working memory rapidly decreases as the items' interactivity increases (Engle and Kane 2004). In other words, information with numerous causal linkages or dependences is difficult to learn because individuals cannot hold both the items and their interactions within working memory. One way to increase working memory processing capacity is by developing schemas, or clusters of information (Clark et al. 2006). Schemas form by connecting pieces of information into a coherent and meaningful structure, sometimes because of considering a single example problem (Ahn et al. 1992). Schemas are held in working memory as a single item. Organizing knowledge in long-term memory into schemas eases the strain on working memory by reducing cognitive load and therefore substantially increases cognitive-processing ability.[2] Improved schema development leads to better-organized knowledge structures, which are typically associated with improved decision-making performance (Sternberg 1997).

Individuals learn by elaborating, or mentally simulating, how new information relates to existing knowledge in long-term memory (Bonner 2007). Mental simulations are cognitive representations of past or potential future events (Kahneman and Tversky 1982). Individuals simulate both by cognitively manipulating how new or existing variables affect an existing mental model and by examining the manipulation's resulting output (Klein 1998; Sanna 2000). Therefore, facilitating the simulation process through mental model development is an important element of learning. Improving auditors' analytical procedures performance requires enhancing their likelihood of accurately simulating expectations of entity-level evidence based on information in their long-term memory.

Using a Systems Perspective to Improve Auditors' Mental Models

To develop expectations about a client's financial position, auditors should construct a sufficiently detailed and accurate mental model of how a client interacts with its business environment (Bell et al. 1997; Bell et al. 2002, 2005). The types of business environments modeled by auditors are complex systems that include features, such as feedback loops and time delays, that can produce nonlinear movement of key system components (Sterman 2000). For example, the time-series movements of entity-level evidence indicators, such as commodity prices and inventory balances, are the culmination of multiple causal effects having different strengths over time. Making accurate inferences and expectations about such indicators requires that decision-makers have mental models that include a coherent sequence of causal relations allowing the auditor to simulate the indicators' movements.

Decision-makers have difficulty learning in, and modeling, complex business environments with nonlinear behavior even when they are presented with all necessary information (Sterman 1989; Sweeny and Sterman 2000). This difficulty results from working memory limitations that prevent individuals from combining and correctly organizing the necessary causal relations caus-

[2] Cognitive load is the burden that performing a task places on a decision-maker's cognitive system, and increases by specific aspects of a task, such as complexity (Paas et al. 2003).

ing such behavior (Sterman 1989). Working memory constraints hinder retrieving and using knowledge in long-term memory while performing mental simulations. Consequently, when placed in an overwhelming business setting, decision-makers often resort to a heuristic-driven reductionist perspective that considers the system components in isolation and bases decisions on more predictable linear relationships (Sterman 2000). Reductionist-based thinking is a common cause of judgment error in complex business settings (Reason 1990; Sterman 2000).

Auditors can reduce reliance on reductionist thinking by taking a systems-based perspective of the client environment (Richmond 1994; Peecher et al. 2007). Bell et al. (1997, 14) describes a systems view as one that "emphasizes basic principles of organization—how the parts are interrelated and coordinated into a unified whole." Such a perspective entails focusing on the causal relationships among system components as well as the nonlinear behavior resulting from these system interactions (Richmond 1994; Sterman 2000).[3] Auditors taking a systems perspective should more coherently organize the environment's causal relations in their mental models, thereby enhancing their ability to develop causal schemas. Therefore, auditors adopting a systems-thinking perspective, and developing schemas of the various environment causal relations, are likely to process more information concurrently in working memory and thereby overcome some of the aforementioned cognitive constraints.[4] These auditors' mental model components should show a logical causal sequence that allows them to better understand the dynamic nonlinear features of a client environment. This improved coherent organization allows the auditor to more easily access knowledge from long-term memory and simulate how entity-level evidence affects the client. Because these auditors can better mentally simulate how economic events flow down to affect the client, they will more likely develop accurate expectations of industry dynamics.

H1: Participants using a systems perspective will develop expectations regarding entity-level evidence that are more accurate than will participants using a reductionist perspective.

The following hypothesis examines the cause of these underlying improvements by investigating whether a systems perspective does in fact reduce auditors' cognitive-processing constraints when performing mental simulations. Prior research shows that prompting auditors to perform simulation-like tasks can improve performance (Earley 2001; Moreno et al. 2007), but this research does not examine how to improve performance when working memory is overwhelmed and increasing cognitive effort is ineffective. As noted earlier, a systems view should help auditors organize information into schemas, thereby expanding working memory capacity. To examine the improvements resulting from expanded working memory, researchers measure the *efficiency* of

---

[3] Systems thinking is a byproduct of system dynamics, which is an interdisciplinary field that uses mathematical computer models to simulate complex quantitative problems that decision-makers cannot perform without such aid (e.g., hurricane forecasting, predicting spread of viruses). System dynamics views the world as a web of interacting complex relationships including feedback loops and time delays. Systems theorists have tried to improve individual decision-making by training people to adopt the same principles that make the system dynamic models successful (i.e., understanding multiple causality, feedback loops, etc.). This training typically involves either building computer simulations or developing skill sets that can recognize dynamic interactions. To date, there is limited laboratory evidence showing either the effectiveness of such techniques or an efficient way to implement such methods.

[4] A system is a "collection of parts that interact to function as a whole" (Bell et al. 1997, 14). As described in this paper, a systems view understands phenomena in terms of these interacting and interdependent parts. In contrast, a reductionist view understands phenomena as isolated individual components. I consider these views as parts of a continuum with system and reductionist endpoints. As individuals construct mental models to guide their viewpoints, they will fall somewhere on this continuum. Information-processing ability is one factor that moderates placement on this continuum and a lack of such ability pushes an individual more toward a reductionist perspective. In this study, I use the term "systems view" as a viewpoint that is relatively more toward the systems endpoint, and "reductionist view" as one that is relatively more toward the reductionist endpoint.

information processing (Paas et al. 2003). Processing efficiency is a byproduct of working-memory capacity and considers both performance and exerted mental effort concurrently. Information processing is more efficient if increases in task performance are higher than expected considering mental effort level, or if mental effort is lower than expected considering performance.

> **H2:** Participants taking a systems perspective will process information in working memory more efficiently when performing mental simulations than will participants taking a reductionist perspective.

Improving Performance: Evaluating Management Representations

During analytical procedures, auditors compare client-provided values, either account balances or other amounts provided in explanations, to quantitative expectations of entity-level evidence (Koonce 1993; Hirst and Koonce 1996). Psychology theory suggests that when a client provides an amount, auditors will simulate a range of potential values to compare against that amount (Mussweiler and Strack 2001; Epley and Gilovich 2006). If the client-provided amount is inconsistent with the auditor's simulated range, then the auditor discounts the representation's credibility. Consequently, accurately simulating this range of potential values increases the likelihood of identifying management representations that are inconsistent with entity-level evidence.

Prior literature examines this expectation violation phenomenon, but these studies either (1) rely on the auditors' prior experience to understand entity-level evidence (Choo and Trotman 1991; Earley 2002), or (2) provide participants with a means to examine *concurrently* entity-level evidence along with client-provided evidence (Earley 2001; Vera-Munoz et al. 2007). To date, no studies separately examine the mental model and expectation development process and how they affect *subsequent* evaluation of client representations. Distinguishing between expectation development and evidence evaluation is important because auditors often rely on their prior knowledge for corroborating client-provided evidence, and extant literature shows that auditors can be influenced if they cannot discredit such evidence (Kennedy 1995; Earley et al. 2008).

The aforementioned theory suggests that an ordinal interaction exists between client analysis perspective and management representation consistency with entity-level evidence. Systems auditors will be better able to mentally simulate a range of potential values for entity evidence and therefore distinguish between management representations that are consistent with such evidence from those that are not. These auditors are more likely to provide lower credibility assessments for inconsistent representations. Furthermore, when management representations are inconsistent with entity evidence, systems auditors should be more likely to recognize the inconsistency and provide lower credibility ratings than reductionist auditors. The following hypothesis sets forth the prediction:

> **H3:** Participants will evaluate the credibility of management representations at the lowest level when they use a systems perspective of entity-level evidence and when they receive management representations that are inconsistent with that evidence; participants will evaluate the credibility of management representations at a higher level under all other combinations of analysis perspective and evidence consistency.

Improving Performance: Updating Expectations of Entity-Level Evidence

In addition to improving processing ability, a coherently organized mental model provides a structure for assimilating and retrieving incoming information (Brewer and Nakamura 1984; Sweller et al. 1998). In an audit context, Bonner et al. (1997) find that relevant knowledge

structures in place prior to experiencing events aid in learning from those events. I extend this result by examining one potential component of these learning improvements, namely that organizing information reduces the constraints on working memory capacity and, hence, improves an individual's understanding of how to accurately integrate new information to revise a knowledge structure.[5] While theoretical frameworks suggest that knowledge structure and working memory interact in the learning process (Kyllonen and Woltz 1989; Bonner 2007), there is little empirical evidence showing how this process occurs. This study tests how organized knowledge improves mental simulation in working memory allowing for improved mental model building.

A better-developed mental model established through a systems-thinking perspective likely improves auditors' learning in a dynamic business setting for two reasons. First, the resulting improved simulation coherence should allow the decision-maker to effectively simulate how the various system components can affect each other in a logical sequence. Second, schemas of the causal interactions will reduce cognitive load and increase the likelihood of understanding how the new information affects the dynamic nonlinearities resulting from component interactions. Understanding the environments' dynamics allows decision-makers to manipulate how the new information affects the essential causal relations within their existing mental models.

Auditors using a better-organized mental model with a systems-based perspective will be more likely to understand how new evidence affects the existing cause-and-effect relationships and resulting nonlinearities within their mental model. This understanding makes these auditors more likely to change their existing inferences regarding such evidence when they receive new, contradictory information (Bell et al. 2005).

> **H4:** When participants receive new entity-level evidence, those who have relatively more coherently organized mental models and take a systems perspective will most accurately update their mental models to modify existing expectations regarding entity-level evidence compared to all other participants.

III. METHOD

Participants

The participants are 61 accounting students from various junior- and senior-level accounting courses at a large state university, randomly assigned to experimental conditions. The theories underlying this study apply to both expert and novice auditors, but for present purposes, using novices has distinct advantages. First, using student participants shows that less experienced auditor subjects garner process improvements through systems-thinking, contrary to recent critical claims that only experienced auditors can perform the complex business environment analysis required by professional standards (AICPA 2006; Curtis and Turley 2007).

Second, this subject pool affords a more powerful test of theory than a professional subject pool (Smith 1989; Friedman and Sunder 1994).[6] Experimental tasks must strain working memory to reveal the differences between conditions (Tindall-Ford et al. 1997). The task I employ is more

5 My predictions suggest that improved working memory processing ability is *one component* of mental model updating. I do not preclude that improved knowledge structures would benefit updating even if one held processing ability constant (e.g., an impoverished model might not have developed the links essential to understanding how the new information would affect the existing model). My experimental design addresses this issue by examining whether task-performance improvements are due to knowledge content differences alone and/or include an aspect of processing ability.

6 Smith (1989) conjectures that the primary purpose of laboratory experiments is to test theory and researchers should design experimental procedures to maximize this goal. Furthermore, Friedman and Sunder (1994, 21) state that experi-

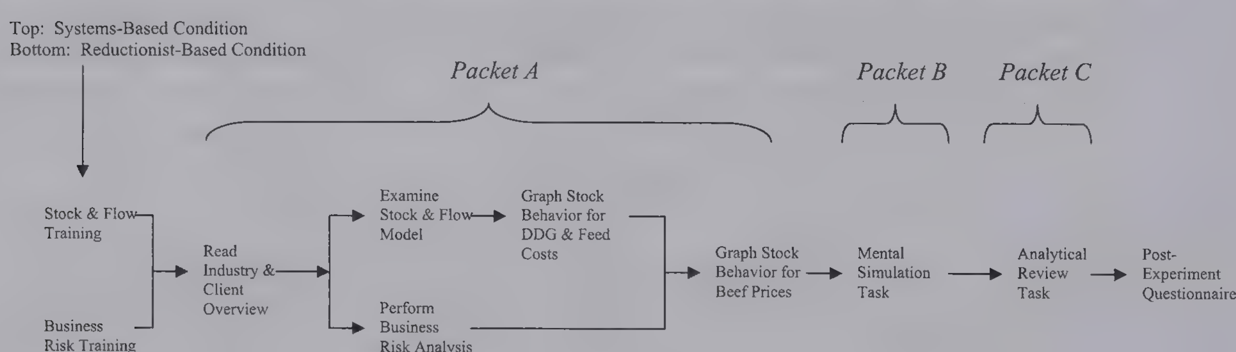
likely to strain working memory, and to do so to a greater degree, for nonprofessional, student subjects than audit professionals because efficient use of working memory resources is a common attribute of expertise (Sternberg 1997).

Also, because my experiment involves training in systems-thinking and the process of learning entity-level evidence, using auditor participants with variable, specialized, pre-developed routines for or disposition toward analyzing such information would not facilitate a clean mental model manipulation. Using student participants with less variable, less specialized, and more malleable knowledge structures, relative to professionals, allows me to effectively control and manipulate mental model development (Libby et al. 2002). Holding constant the participants' pre-existing knowledge structures by using nonprofessional subjects—an approach consistent with previous research examining learning environments (Bonner and Walker 1994; Bonner et al. 1997)—is conducive to identifying how knowledge organization and information-processing ability interact to improve performance.[7]

Task

The instrument contains tasks (see Figure 1) representing analytical procedures performance from initial evidence gathering and mental model building to subsequent management representation evaluation (Koonce 1993; Bell et al. 2005). First, participants receive a tutorial describing how auditors use entity-level information for developing client mental models. Participants then

FIGURE 1
Overview of Experimental Procedures



This figure represents the temporal sequence of the experimental materials that the participants received. Each Packet was taken from the participants before they received the additional items (although they were allowed to use the training tutorial while completing Packet A).

mental design should "sharpen the effects of focus variables and minimize blurring due to nuisance variables." Controlling for pre-existing knowledge by using nonprofessional subjects supports this objective by minimizing any undue effects this knowledge could have on my dependent variables.

7 In addition to these theory-testing advantages, this research pool is also sufficient for my study; auditor participants are unnecessary because no current theoretical or empirical evidence suggests that pre-existing knowledge would interact with the manipulated factors in such a way as to change the *nature* of the independent variables' predicted influence on the dependent measures (Peecher and Solomon 2001). While the pre-existing knowledge structures of professional auditors could reduce the magnitude of the predicted effects by decreasing strain on working memory, there is no reason to suspect that these structures could also cause a predicted directional effect to reverse (i.e., a disordinal interaction).

opened Packet A, which contained materials for a fictitious hamburger food chain.[8] These materials include an industry and company overview characterizing current conditions in the beef industry (see Exhibit 1). The case describes how increased ethanol production causes the price of corn-based feed for cattle to increase, subsequently increasing beef prices. However, another key point in the case is that the use of an ethanol byproduct, called DDG, for cattle feed causes beef producers' feed costs to decline after June of the period under examination because DDG costs only 60 percent of typical corn-based feed. See Figure 2 for a computer simulation of beef price behavior based on the facts in the case. The case states that the company's success ties heavily to the beef industry and instructs participants to develop expectations by inferring the time-series behavior of beef prices for the year under review.

After turning in Packet A, participants applied their existing mental models to perform a mental simulation task in Packet B by writing a paragraph describing how an abrupt change in ethanol prices would affect the company's profits. After turning in Packet B, participants conducted an analytical procedures task in Packet C. This task shows that the company's profitability declined over the past year and provides both a gross profit analysis and an explanation from management targeting rising beef prices as the reason for the decline. The provided explanation is, however, incomplete, as beef prices did not steadily rise during the year, but declined after June because of implementing the DDG byproduct as feed. The participants assessed the credibility of management's explanation of the profit decline and provided their own estimates of year-end beef prices. Finally, participants received an additional paragraph that explained new facts regarding entity-level evidence and provided updated beef-price estimates considering those facts.

Independent Variables

The experiment employs a 2 × 2 between-subjects design with two manipulated independent variables: (1) *industry analysis perspective* (systems-based or reductionist-based) and, (2) *the degree to which management representations regarding year-end beef prices are consistent with industry evidence* (consistent at \$125, or inconsistent at \$165).

The industry perspective variable provides conditions that aid participants in either developing a systems view or resorting to a reductionist view. The systems-based condition provides training and materials helping participants to develop a mental model with a coherent structure of the causal relations in the case materials. Existing research does not provide a systems-based manipulation that actually improves quantifiable aspects of auditors' mental models. Therefore, my systems manipulation uses stocks and flows, which systems theorists state is the most effective way to develop accurate mental models.[9] Business settings are comprised of multiple stocks and flows, and these items should help individuals simulate the nonlinear movement of system components. However, even experienced individuals have difficultly understanding these concepts without extensive training (Sterman 2000). Therefore, the systems participants in my study used methods that are difficult to learn without such training. Consequently, because my study utilizes a brief tutorial, I rely on three commonly used techniques by systems trainers to develop mental

[8] The name of the hamburger food chain used in the experimental materials is suppressed due to the fact, unknown at the time of the experiment, that a real-life hamburger food chain with a similar name exists. That chain operates in a different part of the U.S. than the university where the participants were recruited, so I have no reason to believe that any participants associated the materials with the actual chain.

[9] Stocks are accumulations within a system, while flows represent the stock increases and decreases over time. To illustrate with a commonly used example, the water in a bathtub would represent a stock, whereas the water coming through the faucet and leaving through the drain are flows (Sterman 2000). A system's dynamic properties (feedback loops, time delays) are a byproduct of the interactions among and behavior of stocks and flows.

EXHIBIT 1

Beef Industry and Company Overview

- Beef prices have increased in recent years. The price of beef as of January 1, 2006 was $100 per hundredweight (cwt) of cattle and reached record levels later in the year.
- The recent increase in beef demand is a sharp contrast to prior periods as the demand for beef was on the decline for the previous 20 years. Consumers moved away from beef to alternatives, such as poultry, that are less expensive and considered healthier. For example, in 1975 beef accounted for roughly 49 percent of meat products consumption but by 1997 that share had fallen to 32 percent. This decline is due to the previously mentioned health concerns about beef as well as the increased prices of beef compared to cheaper alternatives such as pork and poultry. In addition, the beef industry has historically lagged behind both the poultry and pork industry in developing pre-packaged, pre-cooked products that allow consumers to reduce cooking time.
- In the past decade, health concerns by customers have made the industry focus on "natural" beef, which is becoming an expectation of a large and growing portion of the population. For example, consumers in some markets are willing to pay as much as 25 percent more for a round roast if it has a label containing the word "natural."
- Profitability in the beef industry is declining due to rising production costs. Unfortunately, the long life cycle of cattle means that beef producers must make production decisions years in advance in order to respond to market conditions. Cattle have the longest life cycle of any livestock produced, since it takes cattle four years to reach maturity. Furthermore, the litter size for cattle averages only one calf per litter.
- Like all commodity markets, the price of cattle is dependent on supply and demand. In livestock markets, however, production cannot take place without a breeding stock. Producers must choose whether to send their stock to market or to keep animals on the farm to increase the future supply. Producers control the supply of breeding stock, and that stock is increased when producers withhold mature animals from market. Producers will increase their desired breeding stock (to increase future supply) when the expected markup ratio is high and will reduce their breeding stock when the markup is low. In the short term, the longer life cycle of beef reduces the producer's ability to increase capacity to deal with rising demand. Industry experts believe it takes between three and four years for cattle producers to fully respond to price increases by increasing production capacity.
- Over the previous 100 years, the price of feed was relatively stable and predictable when adjusted for inflation. In the United States, almost 60 percent of corn production is used as animal feed. Recent trends in the corn industry have caused demand and prices for corn to increase. Ethanol production has increased the demand for corn, which in turn has increased feed costs for cattle producers. The price of corn affects beef and dairy production because added costs affect profit margins of beef producers. The ethanol industry's demand for feed grain will grow exponentially over the next five years increasing from less than 5 percent of supplies a few short years ago to perhaps 75–80 percent in 2009 if the United States produces an 11–12 billion bushel crop. Typically any price changes in feed costs are reflected in beef prices within a week.
- Over the next few years, ethanol prices are expected to continually rise. The price of feed per bushel was $3.50 on January 1, 2006 and changed throughout the year.
- One consideration of the ethanol process is the production of the Dry Distiller Grains (DDG) byproduct. This byproduct costs approximately 60 percent of what regular corn feed costs and can be fed to cattle. The availability of DDG as a source of feed is dependent on the available supply. Ethanol producers did not start stockpiling DDG until the beginning of January and did not have enough supply to sell until approximately June 2006.
- Another potential effect of high corn prices is that ranchers could bring lighter cattle to market. Livestock weights have been 10 to 20 pounds lower than usual the past year. This will affect the quantity of beef on the market, which will mean less top-quality meat on the market, which will drive up beef costs.

Company Overview for 2006

- XYZ Company is a publicly held, regional fast food restaurant chain specializing in hamburgers. The Company has more than 600 locations and is located in ten different states in the southwestern United States. The corporate headquarters is located in Fort Worth, Texas.

(continued on next page)

EXHIBIT 1 (continued)

- XYZ Company is differentiated from its competitors by serving its hamburgers made-to-order. Each burger is cooked immediately after the customer orders, with a typical cooking time between 5–7 minutes. Management believes this approach ensures a quality product and allows the customers to see the difference in service and quality between XYZ Company and its competitors.
- The company takes great pride in the low turnover of its employees. The corporate office does an annual employee survey to get feedback from its restaurant associates. Furthermore, past results have shown that those stores with the lowest turnover margin are the ones with the highest sales. Since 2002, the associate turnover rate has dropped by 92 percent.
- Despite a history of profitability, management is concerned about the rising beef costs. Approximately, 70% of the company's sales are from beef products, which is substantially higher than most of their competitors that introduced chicken products years ago. Management believes these rising costs will reduce profitability, which will hinder their own ability to expand the menu by introducing a line of chicken products.

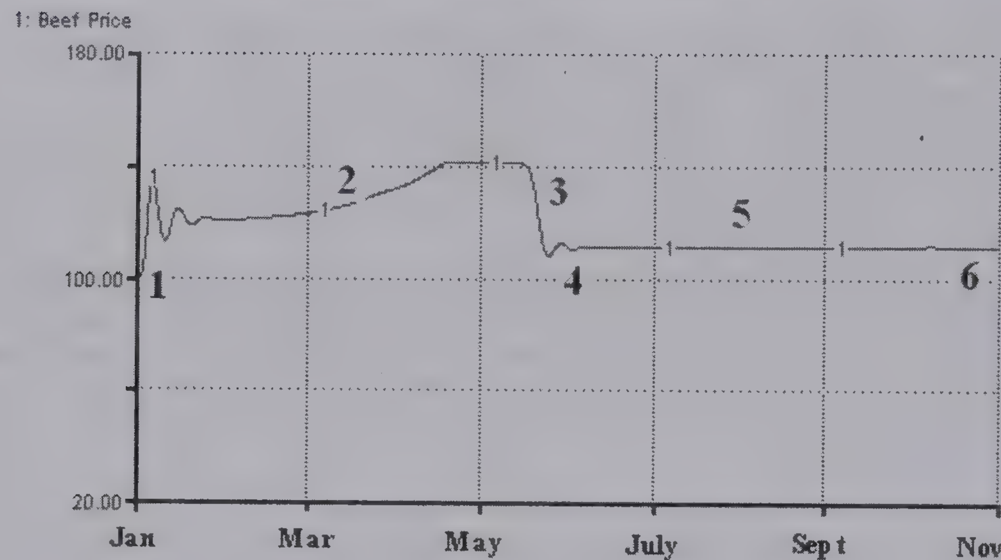
models: (1) diagrams that visually display the causal interactions, (2) active learning techniques to ensure that participants focus on the necessary causal relations, and (3) decomposition of complex causal interactions.

The systems-based participants began the experiment with a tutorial that discusses the nature and interaction of stocks and flows, provides examples, and asks participants to sketch how a stock balance would fluctuate over time. Packet A provides a stock-flow diagram, *based entirely on the industry overview*, showing how the major components of the beef industry interact (see Figure 3). Participants completed an active learning task by writing a sentence describing each interaction within the model. Finally, participants sketched the patterns of behavior for three stocks within the system that have a causal relationship (DDG byproduct, feed prices, and beef prices) using blank graphs similar to that shown in Figure 2.

Participants in the reductionist-based condition used their own intuition to organize a mental model of the case materials. The reductionist participants received the exact same industry and company overview as the systems participants. While they did not have access to the stock/flow tutorial or diagram, I included steps to enhance their understanding of the case materials. First, to ensure that participants understood why constructing a mental model of industry evidence is important and a potentially complex task requiring the assimilation of numerous causal relations, I provided them with a tutorial on business risk auditing, drawing heavily on current auditing standards and methodologies (Bell et al. 1997; Bell et al. 2005; AICPA 2006). Second, I included an active learning task to ensure that the participants studied the industry overview thoroughly enough to construct a mental model of the client environment. Participants identified business risks from the industry overview and described how these risks affect the company. Finally, to reinforce the notion that environment components change over time, participants sketched out the behavior of beef prices over the year.

Both perspective conditions are consistent with current professional standards and business risk auditing methods. The difference in each condition is in the level of aid provided to organize the multiple causal relations in the case. Note that the reductionist participants did not receive the stock-and-flow model, but the information in the industry overview described the facts needed to understand the interactions and nonlinear behavior within the industry *assuming there were no working memory constraints.* The reductionist participants received all the necessary information to understand the nonlinear movement of beef prices. Consequently, I expect them to understand any individual causal relations, but not be able to properly combine them due to working memory

FIGURE 2
Computer Simulation of Beef Prices



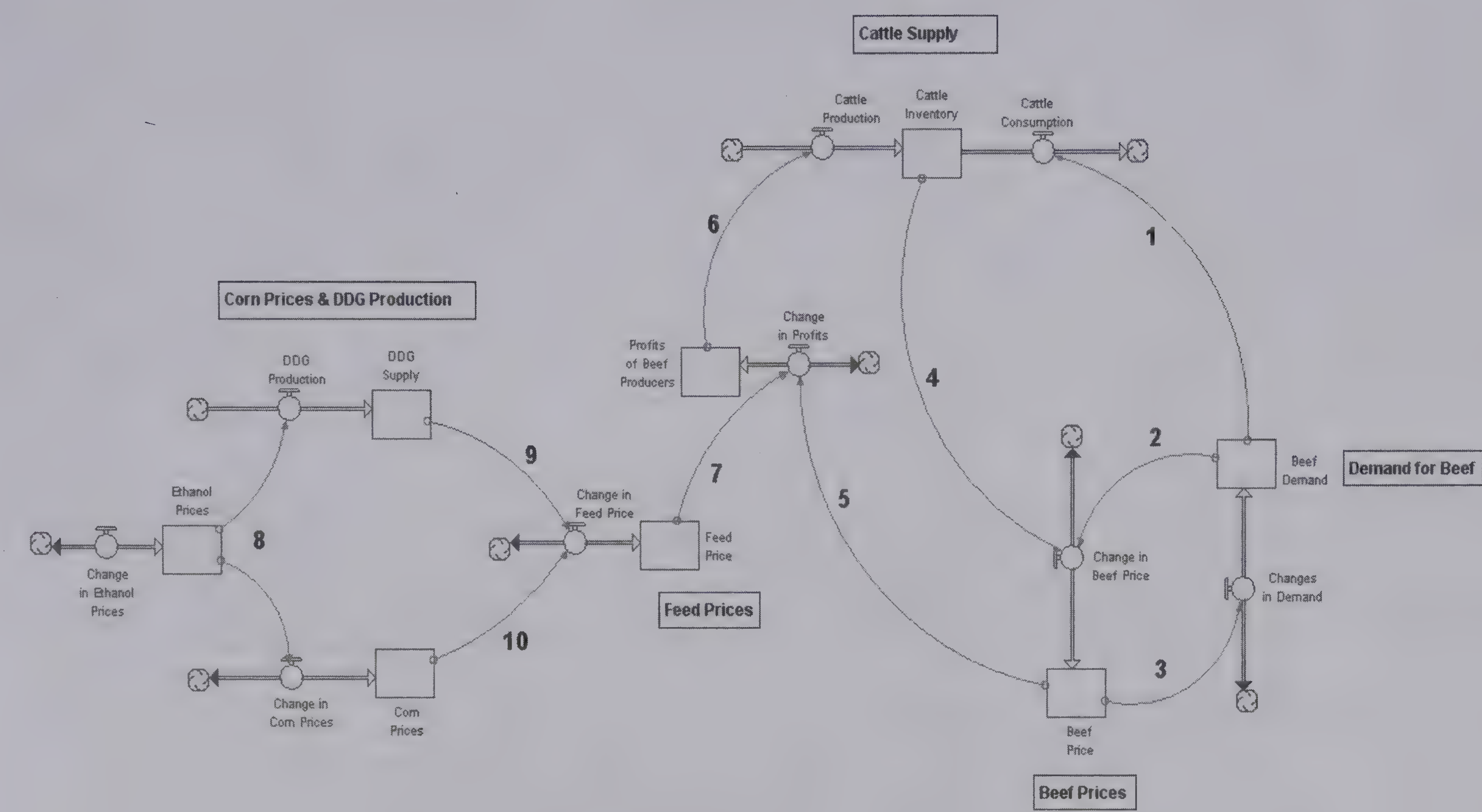
This figure shows a computer simulation of beef prices using the information from the industry overview (Packet A). This simulation model is a supply-and-demand market for commodities based on dynamic models provided in Sterman (2000). The model is coded to move to equilibrium in the absence of external forces as shown by the steady-state periods. The numbers above (1–6) represent features of the simulation used to code the students' graphs of beef price. A point was awarded if the graph exhibited evidence of the above features. The numbers represent the following:

1. The starting point for beef prices is \$100 cwt.
2. Beef prices should rise until approximately June. The rise is due to increasing feed costs as the prices of corn-based feed are rising due to the increased usage of corn for ethanol production.
3. Decline in prices after June as beef producers begin to switch to the ethanol byproduct, DDG, which is 60 percent cheaper than regular corn-based feed.
4. Prices should reach a low point in the periods following June.
5. The market moves back to equilibrium following the decline in feed costs. This can be represented in an increase from the lowest point as long as it is not higher than the pre-DDG period.
6. The final estimate should be lower than the highest point of the graph to reflect the decreased feed costs for beef producers.

limitations. During the post-experiment questionnaire, I verified that participants in both conditions understood the basic causal relations related to beef prices.

My second manipulated independent variable is the consistency level of management representations compared to entity-level evidence. The objective is to provide a representation that is judgmental and requires participants to simulate an outcome from their mental models to evaluate the representation. In the case, management explains that the company's deteriorated profitability stemmed from rising food costs due to increased beef prices. Management states that beef prices had risen to record highs over the past year, and provides a year-end beef price of either \$125 per hundredweight of cattle (cwt) (consistent) or \$165 cwt (inconsistent). Recall, that while participants initially developed expectations, they had a framework (see Figure 2) with a lower boundary

FIGURE 3
Beef Industry Model



This figure represents the model that was provided to students in the Systems-based condition. The participants were instructed to write a sentence describing the effect that each numbered linkage had on the related stocks. The following represents the linkages in the model above. The information below was *not* presented to

the participants, but is included here to help the reader understand the environment:

1. An increase in the demand for beef increases the amount of cattle consumed, and *vice versa*.
2. The demand for beef is one factor that drives beef prices. An increase in demand increases prices, and *vice versa*.
3. Beef prices affect the demand for beef. An increase in prices will decrease demand, and *vice versa*.
4. The supply of cattle directly affects beef prices. An increase in supply will decrease beef prices, and *vice versa*.
5. Beef prices are one component of the profits for beef producers. An increase in price will increase profits, and *vice versa*.
6. The profits of beef producers will affect the amount of cattle produced. As profits rise, there will be more desire to increase production to take advantage of higher profit margins.
7. Feed prices will directly affect the profits of beef producers. An increase in feed prices will reduce profits, and *vice versa*.
8. Ethanol prices affect both the level of DDG byproduct and corn prices. As ethanol prices increase, production will increase and result in an increased level of DDG. Also, increased ethanol production means more corn is needed, which will increase corn prices.
9. The supply of DDG will eventually decrease feed prices since it is less costly than corn-based feed.
10. Corn prices affect feed prices. An increase in corn prices will increase feed prices, and *vice versa*.

of $20 cwt, a midpoint of $100 cwt, and an upper boundary of $180 cwt. I expect participants to use this graphical framework when simulating outcomes for representation evaluation. Therefore, the higher $165 (lower $125) price is less (more) likely to incorporate the DDG-driven decline in the second half of the year. As a result, it is more likely that the $165 ($125) representation is inconsistent (consistent) when considering the entity-level evidence.

Control Variables

Individuals have varying abilities to process and manipulate information in working memory. To control for these variations, participants completed the short-form Raven Advanced Progressive Matrices Test, measuring their ability to manage sets of information in working memory (Raven 1976; Carpenter et al. 1990; Arthur and Day 1994). Additional control variables include self-assessments (using an 11-point scale) of mental effort on all packets, motivation on all packets, understanding of both the industry overview and the tutorial, and minutes spent on each packet.

Dependent Measures

Mental Model Development Measure

The predictions rely on the systems manipulation helping participants to develop more coherently organized mental models. To measure the success of this manipulation, I calculate mental simulation coherence, which captures the causal sequence structure of a mental model. I examine the extent that participants' mental simulation paragraphs have a logical causal chain throughout the paragraph. The paragraphs are coded based on the methodology used by Lesgold et al. (1988) to quantify experts' problem-solving methods because this measure captures participants' ability to perform a mental simulation that ties together components in a coherent progression without disconnections (Klein 1998). Participants list and then visually draw each attribute of the paragraph (i.e., each system component) on a graph. Exhibit 2 provides an example of the simulation paragraphs coding. The coherence metric is a percentage based on the number of system components connected by a single directional causal chain. This method ascribes a higher percentage to those simulations that tie together a larger number of components with a single directional causal chain.

Hypotheses Dependent Measures

To test H1, I evaluate the accuracy of participants' expectations of beef prices. The participants developed beef price expectations by sketching the time-series behavior of the prices for the period under review. I assign a point (maximum of 6 points) if the expectation sketches exhibited the same dynamic properties exhibited in a computer simulation (see Figure 2) using the facts and assumptions presented in the industry overview. Higher scores represent more consistency with the computer simulation.[10,11]

If there are no significant differences between the industry-perspective conditions in regards to the amount of *necessary* information presented, then both conditions should have a basic understanding of the individual causal relations (i.e., the building blocks) driving beef prices. Any mental model improvements should be due to an inability to accurately organize these causal

---

[10] I coded the simulation paragraphs with the assistance of another graduate student, covering all information regarding to which experimental condition the participants' packets were assigned. Both coders worked independently while coding, and mutually resolved any significant differences. The mean correlation index between raters for simulation coherence percentage is 0.85 (Kappa = 0.45, $p < 0.001$).

[11] A graduate student and I individually coded the graph scores and mutually resolved any significant differences. The mean correlation index between raters is 0.97 (Kappa = 0.81, $p < 0.001$).

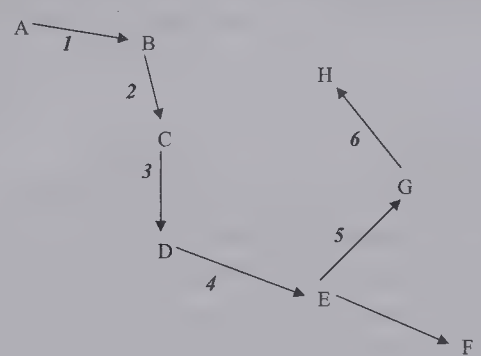
EXHIBIT 2

Example of Mental Simulation Scoring

The items below represent an example of how the mental simulation paragraphs were coded. Each component within the simulation was assigned a letter. Then causal linkages were established based on how the participant described the relationship between components. The coherence measure was calculated based on the number of identified components in the system and the longest directional chain through the system.

Number of Components: The total number of system components identified in the simulation (lettered items below).

Longest Chain: The longest causal chain that exists throughout the simulation (numbered items below).



A. Increase in ethanol prices
B. Increase in corn prices
C. Increase in feed prices
D. Rise in beef prices
E. The company's cost of production
F. The company's use of substitute products
G. Decrease in beef consumption
H. The company's profits
Number of Components: 8
Longest Chain: 6
Simulation Coherence (Longest Chain/# of Components): 0.75

relations, not from a lack of knowledge content of the causal relations. During the post-experimental questionnaire, participants answered multiple-choice questions concerning the relevant causal relations needed for developing accurate beef price expectations.

To test H2, I employ a commonly used workload efficiency measure that establishes a zero-efficiency benchmark and bases performance on a certain effort level (Clark et al. 2006). Participants provided a self-assessment rating of their mental effort level during the simulation paragraph task (using an 11-point scale). Such measures of mental effort are sensitive to variations in exerted effort (see Eggemeier and Wilson [1991] for review). I use simulation coherence as the performance measure in calculating processing efficiency because it represents the participants' level of attainment during their mental simulations (Paas et al. 2003).

To determine processing efficiency, I first recalculate the mental effort rating and task performance measures as z-scores. Then, I calculate processing efficiency as:[12]

$$EFF = \frac{P - R}{\sqrt{2}},$$

where *EFF* is the value of processing efficiency, *P* is the performance z-score, and *R* is the mental effort z-score. Both mental effort (x-axis) and performance (y-axis) are graphed on a Cartesian coordinate grid (see, e.g., Panel C of Table 3). *EFF* represents the perpendicular distance of a point on the axis $P - R$, which is the zero-efficiency line. In other words, on the line $P - R$, mental efficiency is zero ($E = 0$) because performance is at the expected level given the amount of effort exerted. For example, higher mental effort results in higher performance, and vice-versa. If $P - R > 0$, then *EFF* is positive, and if $P - R < 0$, then *EFF* is negative. Any point to the left of the zero-efficiency line ($E = 0$) represents an increase in processing efficiency, and any point to the right represents a decrease in processing efficiency. Experimental conditions are evaluated based on their relative position on the coordinate grid.

To test H3, I evaluate participants' credibility assessments (using an 11-point scale) of management representations regarding the company's decline in profitability. In addition, after evaluating management representations, the participants provided a quantitative range (i.e., confidence interval) for their beliefs regarding year-end beef prices. This measure accomplishes two objectives. First, by analyzing whether the participants' credibility assessments influence their ranges, I support H3 by showing that participants who recognize the inconsistencies incorporate this finding into their ranges rather than anchoring their assessment on management's representation. Second, by showing that the participants' initial ranges include their pre-developed expectations, I establish a benchmark to support my measures for expectation updating (H4), as described in the next section.

To test H4, I present participants with a paragraph including *new information* about the poultry industry. This new information is that poultry, unlike beef, cannot use the DDG byproduct for feed. Correct assimilation of this information into an existing mental model shows that the demand for beef should rise because its substitute product, poultry, will not experience the DDG-related reduction in feed costs and subsequent price decline that beef will. The industry overview (Packet A) states that beef supply cannot increase in the short term. Quantitative ranges regarding beef prices (see H3 dependent measures) should shift upward at year-end, reflecting an increased demand for beef because of higher prices for chicken, a beef substitute. I verified this movement by adding these facts into a computer simulation. Participants provided another quantitative range (confidence interval) for beef prices after considering this new information. I expect systems participants with a coherent mental model to mentally simulate how the new evidence affects the existing model and increase their beef price estimates by shifting both ends of their quantitative range upward. I use an indicator variable to signify whether participants shift their ranges accordingly.[13]

---

[12] This equation is based on the formula for calculating the distance from a particular point to the line ax + by + c = 0 in a coordinate system with coordinates (x, y) (Paas et al. 2003).

[13] I do not examine the magnitude of the changes since there is no information in the case that suggests a specific dollar value change should occur. The dollar change is dependent on any assumptions made by the participant during the initial quantitative range and subsequent mental simulation during updating. Using dollar value of the confidence interval change as a dependent measure would slant the results in favor of those who made the largest dollar changes, whether correct or incorrect.

IV. RESULTS

Descriptive Statistics of Control Variables

The systems participants took more time to complete both the tutorial ($t_{55} = 5.45$, $p = 0.001$) and Packet A ($t_{55} = 2.87$, $p = 0.006$) than did reductionist participants.[14] This outcome is not surprising, considering that the tutorial presented system participants with complex material that is not a part of most educational programs. Consistent with the increase in time, the systems participants' self-assessments of both effort level ($t_{59} = 2.97$, $p = 0.004$) and motivation ($t_{59} = 2.36$, $p = 0.022$) are higher than those of reductionist participants. This outcome is also intuitive because the systems-thinking material requires that participants learn less familiar and more complex concepts than the concepts learned by the reductionist participants. Both groups provide responses indicating that they understood the tutorial and overview, and there is no significant difference between conditions. There also are no significant differences for time and effort between industry perspective conditions for Packets B and C.

Systems View Manipulation Check

I expect that systems participants will generate mental simulations of entity-level evidence that are more coherently organized than those created by reductionist participants. As shown in Table 1, systems participants provide mental simulation paragraphs with a higher percentage of components held together by a single causal chain ($F_{1,57} = 5.39$, $p = 0.012$).[15] This result suggests the systems intervention is effective in aiding how the systems participants organized the entity-level evidence in the case materials.[16]

Hypothesis 1

H1 predicts that systems participants will develop more accurate expectations of beef prices than will reductionist participants. As shown in Table 2, Panel A the systems participants graph scores are significantly higher than those provided by the reductionist participants ($t_{56} = 2.63$, $p = 0.005$).[17] I collapse the individual expectations sketches for each condition in Panel C of Table 2, which shows that systems participants better understand the nonlinear behavior of beef prices and anticipate the decrease in prices after June. A regression analysis (untabulated) reveals that a systems perspective significantly predicts lower beef price estimates at year-end ($t_{56} = -4.50$, $p < 0.001$). These results support H1.

---

[14] The systems (reductionist) participants took 13.53 (8.23) and 20.21 (15.23) minutes to complete the tutorial and packet A, respectively. Using time as a control variable does not qualitatively change the reported results in this study.

[15] Further analysis of the simulation data (not tabulated) shows that self-assessed motivation increases the number of components listed ($t_{58} = 2.28$, $p = 0.013$), the number of causal links between components ($t_{58} = 2.03$, $p = 0.023$), and the length of the longest causal chain ($t_{58} = 1.80$, $p = 0.038$). However, motivation does *not* affect participants' ability to tie all of the relevant components together into a coherent simulation ($t_{58} = 0.58$, $p = 0.283$). This result supports the notion that constraints in working memory capacity are the underlying cause of the systems participants' improved simulations.

[16] Because coherence is a ratio, multiple sets of numbers could provide the same ratio even though they are tying together a different number of components (e.g., 3/4 and 6/8 will both equal 0.75). Since simulations with fewer components could be tied together with relatively little cognitive effort, I ran an additional analysis that excluded any observations with less than seven components (approximately the 25th percentile). When these items are excluded, the systems participants' scores are still significantly higher ($F_{1,48} = 6.00$, $p = 0.009$).

[17] Because effort level was significantly higher for the systems-based participants, the interaction between effort and industry analysis perspective is included in the model. I tested the graph scores for multicollinearity, because there is an expected level of dependency between the industry perspective condition and the self-assessments of effort level. I correct for the multicollinearity using the partial orthogonalization method. All remaining analyses with multicollinearity receive similar corrections. Due to the dependent nature of the variables, I also performed a path analysis, finding quantitatively similar results.

TABLE 1

Manipulation Check: Mental Simulation Coherence

Panel A: Mean (Standard Deviation) Simulation Coherence Scores

Dependent Measure	**Industry Analysis Perspective**	
	Reductionist	**Systems**
Simulation Coherence Scores	0.60	0.70
	(0.18)	(0.18)
	n = 30	n = 31
Components used to calculate Simulation Coherence		
Number of Components	8.95	8.18
	(1.90)	(2.30)
Longest Chain	5.21	5.76
	(1.76)	(2.31)

Panel B: ANCOVA Analysis for Mental Simulation Coherence Scores

Variable	$\mathbf{F_{1,57}}$	**p-value**
Industry Analysis Perspective	5.39	0.012
Effort (Packet B)	0.07	0.396
Raven score	0.02	0.440

Indicator (0/1) coding was used for Industry Analysis Perspective in this table (systems-based perspective = 1; reductionist-based = 0).

All p-values are one-tailed. See Exhibit 2 for coding example.

Variable Definitions:
Simulation Coherence Scores = (Longest Chain/Number of Components);
Number of Components = number of individual components identified in the simulation paragraph; and
Longest Chain = longest causal chain that exists throughout the simulation paragraph.

To determine if both conditions understand the basic knowledge content necessary to estimate beef prices, during the post-experiment questionnaire, participants answered two multiple-choice questions essential to recognizing the nonlinear behavior of beef prices. The questions concern how the DDG byproduct affects feed prices (first question), and how feed prices affect beef prices (second question). Approximately 71 percent of the systems participants correctly answer the first question compared to approximately 76 percent of the reductionist participants ($t_{58} = -0.42$, $p > 0.500$). Approximately 97 percent of both conditions answer the second question correctly ($t_{58} = 0.05$, $p > 0.500$). These results imply that, when prompted, the reductionist participants possess knowledge content of the general causal relations driving beef prices but are unable to incorporate this knowledge properly into their expectations. Such a result is consistent with existing research, finding that decision-makers typically can identify and understand individual cues but typically cannot combine them properly when working memory is overwhelmed (Sterman 1989; Diehl and Sterman 1995; Einhorn 2000). Furthermore, this result lends support for my prediction that the systems view improvements are from knowledge organization differences, not entirely from differences in essential knowledge content between the conditions.

Hypothesis 2

In H2, I predict that participants with a systems perspective will exhibit more efficient working memory processing when performing a mental simulation than will reductionist participants.

TABLE 2

H1: Graph Scores for Beef Price Expectations

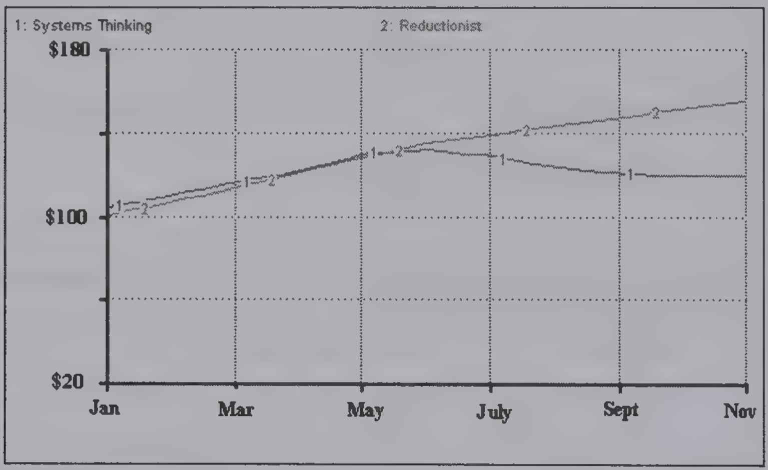
Panel A: Mean Graph Scores (Standard Deviation) of Beef Price Graph Scores

	Industry Analysis Guidance	
Dependent Measure	**Reductionist**	**Systems**
Graph Score (max of six)	2.35	3.25
	(1.02)	(1.32)
	n = 30	n = 31

Panel B: Regression Results for Graph Scores

Variable	β	SE	t	p-value
Constant	1.23	0.88	1.39	0.085
Industry Analysis Perspective	0.87	0.33	2.63	0.005
Effort (Packet A)	0.08	0.12	0.64	0.261
Effort * Industry Analysis Perspective	−0.08	0.24	−0.34	0.368
Raven score	0.08	0.06	1.20	0.117

Panel C: Graphical Representation of Expectations[a]



Indicator (0/1) coding was used for Industry Analysis Perspective in this table (systems-based perspective = 1; reductionist-based = 0).

All p-values are one-tailed.

[a] The graph was calculated by collapsing the participant's beef price graphs for each condition. The resulting average values were input into a systems software program to show the discrepancies between the expectations for each condition.

Variable Definitions:

Beef Price Graph Scores = how many points the participants' beef price expectations received when compared to a computer simulation of beef prices. See Figure 2 for the legend of the computer simulation of beef prices. A point was given if a participant's expectation graph exhibited the properties shown in Figure 2;

Effort = self-assessed mental effort level exerted in Packet A using an 11-point scale; and

Raven = total number of correct questions on the Raven's Progressive Matrices short-form test.

The participants' self-assessed understanding of both the client overview and the tutorial serve as control variables because an understanding of these items is essential for constructing an accurate simulation. Furthermore, because the Raven score controls for innate working memory processing ability, including this score in my analysis highlights any processing improvements stemming from mental model organization. The results in Table 3, Panel B show that the systems participants have more efficient processing when constructing a coherent mental simulation ($F_{1,55} = 4.60$, $p = 0.018$). Panel C provides a graphical presentation and shows that the systems-based efficiency score is above the zero-efficiency benchmark ($E = 0$) line, whereas the reductionist-based score is below the line. These results support H2 and are consistent with prior studies showing that increased use of causal schemas reduces working memory constraints and improves processing efficiency (Paas et al. 2003; Clark et al. 2006).

Hypothesis 3

H3 predicts that systems participants receiving inconsistent representations evaluate those representations as less credible than participants in the other three conditions. Table 4, Panel A provides descriptive statistics and shows a divergence of credibility assessments between the two industry analysis conditions. Panel B presents the conventional ANCOVA results controlling for effort level and Raven score. Because H3 predicts an ordinal interaction, contrast coding is the most appropriate method of testing the hypothesis (Buckless and Ravenscroft 1990). Contrast weights are +3 in the systems/inconsistent condition and −1 in the other three conditions. Panel C shows the planned contrast is significant since systems participants receiving inconsistent management representations provide the lowest credibility ratings ($F_{1,53} = 7.12$, $p = 0.005$). Follow-up tests of the simple effects show that systems auditors are able to distinguish between inconsistent and consistent representations ($F_{1,53} = 4.69$, $p = 0.017$), and are more likely than reductionist auditors to recognize inconsistent representations ($F_{1,53} = 6.62$, $p = 0.001$). Reductionist participants do not distinguish between evidence consistency, suggesting that they have difficulty simulating how the explanation compares to their existing knowledge. These results support H3.

H3 implies that systems participants will fixate their judgments less on management representations, and instead rely more by their own pre-developed expectations. I perform a regression analysis (untabulated) examining how the participants' credibility assessments affect their beef-price quantitative ranges. When the materials present inconsistent representations, the systems participants' credibility assessments are a significant predictor of the upper bound ($t_{48} = 2.26$, $p = 0.014$). This result suggests that those participants recognizing the inconsistency between their pre-developed expectations and management representations are less likely to anchor their inferences about beef prices on management representations (i.e., they are more likely to shift their confidence intervals downward in relation to the inconsistent management representation). By contrast, the reductionist participants' credibility assessments do not have a consistently significant effect on the parameters of the range. These results further support H3.

Hypothesis 4

H4 predicts that participants' most accurate updating of their mental models upon receiving new entity-level evidence will occur when they have a coherently organized mental model and when they take a systems perspective of the industry. Table 5, Panel A shows that the interaction of simulation coherence and systems-based condition significantly predicts whether participants

TABLE 3

H2: Processing Efficiency Results

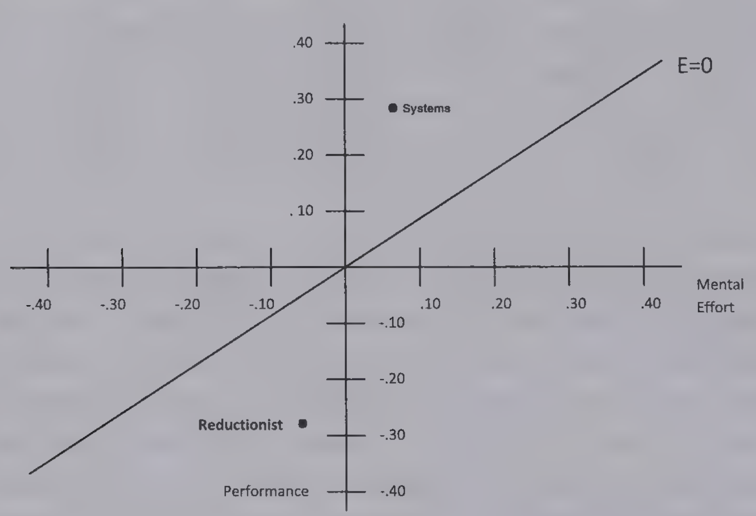
Panel A: Mean (Standard Deviation) Processing Efficiency Scores[a]

	Industry Analysis Guidance	
Dependent Measure	**Reductionist**	**Systems**
Processing Efficiency Score	−0.17 (0.95) n = 30	0.17 (1.05) n = 31
Components of Processing Efficiency Score		
Performance Z-score	−0.29	0.29
Mental Effort Z-score	−0.50	0.50

Panel B: ANCOVA Analysis for Mental Simulation Coherence

Variable	$F_{1,55}$	p-value
Industry Analysis Perspective	4.60	0.018
Understanding—ST Tutorial	1.43	0.119
Understanding—BR Tutorial	1.15	0.144
Understanding—Overview	0.28	0.298
Raven	0.11	0.373

Panel C: Graphical Representation of Results



Indicator (0/1) coding was used for Industry Analysis Perspective in this table (systems-based perspective = 1; reductionist-based = 0).

All p-values are one-tailed.

[a] The Processing Efficiency Score is the relative processing efficiency based on both performance and effort. In this calculation, the simulation coherence score (see Table 2) was used for the performance measures. Effort was the amount of self-assessed mental effort level exerted in Packet B.

Variable Definitions:

Processing Efficiency Score = [(Performance Z-score − Effort Z-score)/Square Root of 2];

Performance Z-score = calculated Z-score based on the simulation coherence score;

Mental Effort Z-score = calculated Z-score based on the self assessment of effort in Packet B;

Understanding—ST Tutorial = self-assessed understanding of the stock and flow tutorial given in the systems thinking condition using an 11-point scale;

(continued on next page)

Understanding—BR Tutorial = self-assessed understanding of the business risk tutorial given in the reductionist condition using an 11-point scale; and
Understanding—Overview = self-assessed understanding of the industry and company overview given to all participants using an 11-point scale.

increase both ends of the range (Wald $\chi^2_{1,52} = 3.12$, $p = 0.039$).[18] I verify the robustness of these results by finding that the processing efficiency measure from H2 (see Panel B of Table 5) yields consistent results (Wald $\chi^2_{1,52} = 4.36$, $p = 0.019$). These results support H4.[19]

In sum, the results of this study show that a systems view improves information-processing ability and therefore the likelihood of acquiring and updating both knowledge content and structure. Most accounting studies assume that the effect of processing ability on knowledge acquisition is relatively static and does not change over the short-run (Libby 1995; Bonner 2007), but studies outside of accounting call for research that considers the implications of more dynamic associations between knowledge and ability (Ceci et al. 2003). My results suggest that a more dynamic relationship exists in the short-term between one component of knowledge, structure, and one component of information-processing ability, working memory capacity. This reinforcing relationship between knowledge organization and processing ability is significant because ignoring it fails to account for substantial variation in task performance (e.g., understanding and modeling complex business environments).

V. CONCLUDING REMARKS

This study examines whether a systems-based perspective of a client economic environment improves novice auditors' knowledge acquisition and subsequent analytical procedures performance. Auditing theorists suggest that training auditors to use systems-thinking skills can improve how they learn and use complex entity-level evidence that would normally overwhelm auditors (Bell et al. 1997; Bell et al. 2002). There is little empirical evidence showing the effectiveness of system approaches, and researchers and practitioners have struggled to find a successful and efficient way to implement systems skills due to the extensive training that is typically necessary.

I conduct a laboratory experiment in which novice participants learn about complex entity-level evidence for use in analytical procedures. I find that participants taking a systems view of the client environment develop more accurate and coherent mental models than participants using a reductionist perspective. In turn, systems participants are more likely to identify management

[18] In Panel A, the Raven score significantly reduces the chances of increasing the confidence interval (Wald $\chi^2_{1,52} = 5.97$, $p = 0.007$), suggesting that participants with higher innate processing ability are more likely to *incorrectly* reduce the confidence interval. A further examination (untabulated) finds that the Raven score is significant for the reductionist-based condition (Wald $\chi^2_{1,52} = 3.58$, $p = 0.029$) but not for the systems-based condition (Wald $\chi^2_{1,52} = 1.22$; $p = 0.135$). To lend some insight into this result (untabulated), I find that the Raven score increases reductionist participants' self-assessed confidence level in evaluating beef prices ($t_{55} = 1.46$, $p = 0.075$) but decreases the systems participants' confidence level ($t_{55} = -1.44$, $p = 0.078$). This distinction implies that those with higher "fluid intelligence" levels, but with an impoverished mental model, are more likely to be overconfident in their ability to understand and utilize industry evidence. Conversely, those with better-organized knowledge structures recognize the complexity of the problem and are not overconfident in their ability to interpret it. This result supports the findings of Choy and King (2005), who suggest that decision-makers over-estimate their own innate abilities when solving dynamic problems.

[19] I also examine other attributes of the updated beef price confidence intervals as dependent measures. The interaction of simulation coherence and systems thinking is either significant or marginally significant when using a separate indicator variable for an increase in the lower bound (Wald $\chi^2_{1,52} = 3.44$, $p = 0.032$) and the upper bound (Wald $\chi^2_{1,52} = 2.43$, $p = 0.059$) of the confidence interval. With these two measures, neither simulation coherence nor systems thinking are, in isolation, significant. Finally, I examine whether there was a tendency to incorrectly *reduce* the confidence intervals, finding no significant results for any condition.

TABLE 4

H3: Credibility Assessment of Management Representation

Panel A: Mean (Standard Deviation) of Credibility Assessments

	Industry Analysis Perspective		
	Reductionist	**Systems**	**Totals**
Consistent—$125	7.07 (1.33) n = 14	7.57 (2.06) n = 14	7.32 (1.72) n = 28
Inconsistent—$165	7.25 (1.49) n = 16	6.17 (1.53) n = 15	6.73 (1.67) n = 31
Totals	7.16 (1.49) n = 30	6.84 (1.91) n = 29	7.01 (1.70) n = 59

Panel B: Conventional ANCOVA Analysis

Variable	**Sum of Squares**	**df**	**Mean Square**	**F**	**p-value**
Industry Analysis Perspective (IAP)	2.307	1	2.307	0.89	0.175
Representation consistency (RC)	6.627	1	6.627	2.54	0.058
IAP * RC	11.203	1	11.203	4.30	0.022
Effort (Packet C)	13.959	1	13.959	5.36	0.012
Raven score	0.000	1	0.006	0.00	0.497
Error	138.031	53	2.604		

Panel C: Contrast Analysis and Follow-Up Simple Effects Test

Contrast	**DF**	**F**	**p-value**
Summary Hypothesis: Combination of inconsistency and systems-based perspective difference from other three conditions.	1,53	7.12	0.005
Follow-Up Simple Effect Tests			
Simple effect of Industry Analysis Perspective given inconsistent management explanation	1,53	6.62	0.001
Simple effect of consistent explanation versus inconsistent explanation given systems-based perspective	1,53	4.69	0.017

Indicator (0/1) coding was used for Industry Analysis Perspective in this table (systems-based = 1; reductionist-based = 0). Indicator (0/1) coding was also used for Representation Consistency (inconsistent = 1; consistent = 0).

All p-values are one-tailed equivalent.

Variable Definitions:
Effort = the self-assessed mental effort level exerted in Packet C using an 11–point scale;
Raven = the total number of correct questions on the Raven's Progressive Matrices short form test.

TABLE 5

H4: Updating of Mental Model

$$\text{Increase Confidence Interval} = \alpha + \beta_1(\text{Anchor}) + \beta_2(\text{Industry Analysis Perspective}) + \beta_3(\text{Simulation Coherence}) + \beta_4(\text{Simulation Coherence} * \text{Industry Analysis Perspective}) + \beta_5(\text{Raven}) + \beta_6(\text{Effort}) + \varepsilon.$$

Panel A: Regression Results for Simulation Coherence

Variable	β	SE	Wald χ^2	p-value
Constant	2.00	5.83	0.24	0.310
Anchor	−0.03	0.03	1.13	0.145
Industry Analysis Perspective (IAP)	−2.56	2.60	0.97	0.162
Simulation Coherence	4.07	5.33	0.58	0.223
Simulation Coherence * IAP	18.70	10.59	3.12	0.039
Raven	−0.67	0.27	5.97	0.007
Effort (Packet C)	0.31	0.44	0.52	0.236

Panel B: Regression Results for Processing Efficiency

Variable	β	SE	Wald χ^2	p-value
Constant	4.62	5.14	0.81	0.185
Anchor	−0.02	0.03	0.52	0.235
Industry Analysis Perspective (IAP)	−1.11	1.82	0.37	0.272
Processing Efficiency	0.47	0.92	0.25	0.307
Processing Efficiency * IAP	3.62	1.74	4.36	0.019
Raven	−0.72	0.27	6.01	0.007
Effort (Packet C)	0.19	0.47	0.15	0.351

All p-values are one-tailed.

Variable Definitions:

Increase Confidence Interval = indicator variable that represents whether the participant increased both ends of their confidence intervals; coded 1 if both bounds are increased, 0 otherwise;

Processing Efficiency = processing efficiency score used to test H2;

Effort = self-assessed mental effort level exerted in Packet C using an 11-point scale; and

Raven = total number of correct questions on the Raven's Progressive Matrices short-form test;

Anchor = the beef price amount provided in management's explanation.

representations that are inconsistent with their expectations of entity-level evidence. In addition, the better-organized mental models of the systems participants provide more efficient use of working memory capacity, which aids the accurate updating of their expectations when in the presence of new industry-level evidence. These results are consistent with educational psychology findings that individuals can learn more effectively when they construct the incoming evidence as causal schematic structures (Clark et al. 2006). This study provides insight into how knowledge structure and processing ability interact to enhance learning, shortening the path to building auditors' expert-consistent performance when performing complex audit tasks (Nelson and Tan 2005).

Professional standards require auditors to understand complex entity-level evidence (AICPA 2006). Therefore, this study highlights a potential disconnect between what standards now require of auditors and what they are cognitively capable of doing. This disconnect suggests that new systems-based training methods could help auditors to learn better in the field and develop the skill sets needed to understand complex environments (PwC 2003; Peecher et al. 2007). Leaders in systems-thinking suggest that quantitative model building is the most effective way to fully learn and understand dynamically complex systems (Hannon and Ruth 1997; Sterman 2000). Future research could examine if new tools and decision aids, such as dynamic modeling, can help auditors accelerate learning about complex client environments.

Limitations

A potential concern with this study is whether my systems-based manipulation could be cost prohibitive in an actual audit because it would require the provision of pre-developed diagrams for each client. Recall that the purpose of this study is to examine whether and how systems techniques can improve audit quality, not to provide a specific methodology for firms. The systems manipulation is simply a tool to examine the improvements of a systems-consistent mental model. Therefore, while the diagram utilized in this study was necessary to help participants develop a systems view, I do not expect such a pre-developed diagram to be necessary in an actual audit. Systems advocates suggest the type of information provided in the systems manipulation could be either cognitively or formally built by someone properly trained in such skills (Hannon and Ruth 1997; Sterman 2000; Bell et al. 2002). Whether auditors can effectively and efficiently actually build these systems-consistent mental models remains a question for future research.

REFERENCES

Ahn, W., W. F. Brewer, and R. J. Mooney. 1992. Schema acquisition from a single example. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18 (2): 391–412.

American Institute of Certified Public Accountants (AICPA). 2006. *Understanding the Entity and Its Environment and Assessing the Risks of Material Misstatement.* Statement on Auditing Standard No. 109. New York, NY: AICPA.

Arthur, W., Jr., and D. V. Day. 1994. Development of a short form for the Raven Advanced Progressive Matrices Test. *Educational and Psychological Measurement* 54 (2): 394–403.

Baddeley, A. 1998. *Human Memory: Theory and Practice*. Boston, MA: Allyn and Bacon.

Bell, T. B., F. O. Marrs, I. Solomon, and H. Thomas. 1997. *Auditing Organizations Through A Strategic Systems Lens: The KPMG Business Measurement Process*. Montvale, NJ: KPMG LLP.

———, M. E. Peecher, and I. Solomon. 2002. *The Strategic-Systems Approach to Auditing: Cases in Strategic-Systems Auditing*. Montvale, NJ: KPMG LLP.

———, ———, and ———. 2005. *The 21st Century Public Company Audit: Conceptual Elements of KPMG's Global Audit Methodology*. Montvale, NJ: KPMG LLP.

Bonner, S. E., and P. L. Walker. 1994. The effects of instruction and experience on the acquisition of auditing knowledge. *The Accounting Review* 69 (1): 157–178.

———, R. Libby, and M. W. Nelson. 1997. Audit category knowledge as a precondition to learning from experience. *Accounting, Organizations and Society* 22 (5): 387–410.

——— 2007. *Judgment and Decision Making in Accounting*. Upper Saddle River, NJ: Pearson Prentice Hall.

Borthick, A. F., M. B. Curtis, and R. S. Siram. 2006. Accelerating the acquisition of knowledge structure to improve performance in internal control reviews. *Accounting, Organizations and Society* 31: 323–342.

Brewer, W. F., and G. V. Nakamura. 1984. The nature and functions of schemas. In *Handbook of Social Cognition*, edited by R. S. Wyer, Jr., and T. K. Srull, Vol. 1: 119–160. Hillsdale, NJ: Lawrence Erlbaum Associates.

Buckless, F. A., and S. P. Ravenscroft. 1990. Contrast coding: A refinement of ANOVA in behavioral analysis. *The Accounting Review* 65 (4): 933–945.

Carpenter, P. A., M. A. Just, and P. Shell. 1990. What intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review* 97 (3): 404–431.

Ceci, S. J., S. M. Barnett, and T. Kanaya. 2003. Developing childhood proclivities into adult competencies: The overlooked multiplier effect. In *The Psychology of Abilities, Competencies, and Expertise*, edited by E. L. Grigorenko, and R. J. Sternberg, 70–91. London, U.K.: Cambridge University Press.

Choo, F., and K. T. Trotman. 1991. The relationship between knowledge structure and judgments for experienced and inexperienced auditors. *The Accounting Review* 66 (3): 464–485.

Choy, A. K., and R. R. King. 2005. An experimental investigation of approaches to audit decision making: An evaluation using systems-mediated mental models. *Contemporary Accounting Research* 22 (2): 311–350.

Clark, R., F. Nguyen, and J. Sweller. 2006. *Efficiency in Learning: Evidence-Based Guidelines to Manage Cognitive Load*. San Francisco, CA: Pfeiffer.

Conway, A. R. A., M. J. Kane, and R. W. Engle. 2003. Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences* 7 (12): 547–552.

Curtis, E., and S. Turley. 2007. The business risk audit—A longitudinal case study of an audit engagement. *Accounting, Organizations and Society* 32 (4–5): 439–461.

Diehl, D., and J. D. Sterman. 1995. Effects of feedback complexity on dynamic decision making. *Organizational Behavior and Human Decision Processes* 62 (2): 198–215.

Earley, C. E. 2001. Knowledge acquisition in auditing: Training novice auditors to recognize cue relationships in real estate valuation. *The Accounting Review* 76 (1): 81–97.

——— 2002. The differential use of information by experienced and novice auditors in the performance of ill-structured audit tasks. *Contemporary Accounting Research* 19 (4): 595–614.

———, V. B. Hoffman, and J. J. Joe. 2008. Reducing management's influence on auditors' judgments: An experimental investigation of SOX 404 assessments. *The Accounting Review* 83 (6): 1461–1485.

Eggemeier, F. T., and G. F. Wilson. 1991. Performance-based and subjective assessment of workload in multi-task environments. In *Multiple-Task Performance*, edited by D. L. Damos, 217–278. Bristol, PA: Taylor & Francis Ltd.

Einhorn, H. J. 2000. Expert judgment: Some necessary conditions and an example. In *Judgment and Decision Making: An Interdisciplinary Reader*, edited by T. Connolly, H. R. Arkes, and K. R. Hammond, 324–335. London, U.K.: Cambridge University Press.

Engle, R. W., and M. J. Kane. 2003. Executive attention, working memory, capacity, and a two-factor theory of cognitive control. *Psychology of Learning and Motivation* 44: 145–199.

Epley, N., and T. Gilovich. 2006. The anchoring and adjustment heuristic: Why the adjustments are insufficient. *Psychological Science* 17 (4): 311–318.

Friedman, D., and S. Sunder. 1994. *Experimental Methods: A Primer for Economists*. New York, NY and London, U.K.: Cambridge University Press.

Hannon, B., and M. Ruth. 1997. *Modeling Dynamic Biological Systems*. New York, NY: Springer.

Hecht, G. 2010. The influence of auditor's understanding of circular causality on unintended consequence identification and control risk assessments. Working paper, Emory University.

Hirst, D. E., and L. Koonce. 1996. Audit analytical procedures: A field investigation. *Contemporary Accounting Research* 13 (2): 457–486.

Kahneman, D., and A. Tversky. 1982. The simulation heuristic. In *Judgment and Uncertainty: Heuristics and*

Biases, edited by D. Kahneman, P. Slovic, and A. Tversky, 201–208. London, U.K.: Cambridge University Press.

Kennedy, J. 1995. Debiasing the curse of knowledge in audit judgment. *The Accounting Review* 70 (2): 249–272.

Klein, G. 1998. *Sources of Power: How People Make Decisions*. Cambridge, MA: The MIT Press.

Knechel, W. R. 2007. The business risk audit: Origins, obstacles, and opportunities. *Accounting, Organizations and Society* 32 (4-5): 383–408.

Koonce, L. 1993. A cognitive characterization of audit analytical review. *Auditing: A Journal of Practice & Theory* 12 (Supplement): 57–76.

Kyllonen, P., and D. Woltz. 1989. The role of cognitive factors in the acquisition of a cognitive skill. In *Abilities, Motivation, & Methodology: The Minnesota Symposium on Learning and Individual Differences*, edited by R. Kanfer, P. Ackerman, and R. Cudeck, 239–280. Hillsdale, NJ: Lawrence Erlbaum Associates.

Lesgold, A., R. Glaser, H. Rubinson, D. Klopfer, P. Feltovich, and Y. Wang. 1988. Expertise in a complex skill: Diagnosing x-ray pictures. In *The Nature of Expertise*, edited by M. T. H. Chi, R. Glaser, and M. J. Farr, 311–342. Hillsdale, NJ: Lawrence Erlbaum Associates.

Libby, R. 1995. The role of knowledge and memory in audit judgment. In *Judgment and Decision-Making Research in Accounting and Auditing*, edited by R. H. Ashton, and A. H. Ashton, 176–206. London, U.K.: Cambridge University Press.

———, R. Bloomfield, and M. W. Nelson. 2002. Experimental research in financial accounting. *Accounting, Organizations and Society* 27: 775–810.

Moreno, K. K., S. Bhattacharjee, and D. M. Brandon. 2007. The effectiveness of alternative training techniques on analytical procedures performance. *Contemporary Accounting Research* 24 (3): 983–1014.

Mussweiler, T., and F. Strack. 2001. The semantics of anchoring. *Organizational Behavior and Human Decision Processes* 86 (2): 234–255.

Nelson, M., and H. T. Tan. 2005. Judgment and decision making research in auditing: A task, person, and interpersonal interaction perspective. *Auditing: A Journal of Practice & Theory* 24 (Supplement): 41–71.

O'Donnell, E., and J. Schultz. 2005. The halo effect in business risk audits: Can strategic risk assessment bias auditor judgment about accounting details? *The Accounting Review* 80 (3): 921–938.

Paas, F., J. E. Touvinen, H. Tabbers, and P. W. M. Van Gerven. 2003. Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist* 38 (1): 63–71.

Peecher, M. E., and I. Solomon. 2001. Theory and experimentation in studies of audit judgments and decisions: Avoiding common research traps. *International Journal of Auditing* 5: 193–203.

———, R. Schwartz, and I. Solomon. 2007. It's all about audit quality: Perspectives on strategic-systems auditing. *Accounting, Organizations and Society* 32 (4–5): 463–485.

PricewaterhouseCoopers (PwC). 2003. *Educating for the Public Trust: The PricewaterhouseCoopers Position on Accounting Education*. New York, NY: PricewaterhouseCoopers.

Raven, J. C. 1976. *Advanced Progressive Matrices Set II*. San Antonio, TX: Harcourt Assessments, Inc. Adapted and reproduced with permission. All rights reserved. "*Raven's Progressive Matrices*" is a trademark of Harcourt Assessment, Inc. registered in the United States of America and and/or other jurisdictions.

Reason, J. 1990. *Human Error*. London, U.K.: Cambridge University Press.

Richmond, B. 1994. Systems thinking/system dynamics: Let's just get on with it. *System Dynamics Review* 10 (2–3): 135–157.

Rickheit, G., and L. Sichelschmidt. 1999. Mental models: Some answers, some questions, some suggestions. In *Mental Models in Discourse Processing and Reasoning*, edited by G. Rickheit, and C. Habel, 9–40. Amsterdam, The Netherlands: Elsevier Science.

Sanna, L. J. 2000. Mental simulation, affect, and personality: A conceptual framework. *Current Directions in Psychological Science* 9 (5): 168–173.

Securities and Exchange Commission (SEC). 2003. *Report Pursuant to the Section 704 of the Sarbanes-Oxley Act of 2002*. Washington, D.C.: SEC.

Smith, V. L. 1989. Theory, experiment and economics. *The Journal of Economic Perspectives* 3 (1): 151–169.

Sterman, J. D. 1989. Misperceptions of feedback in dynamic decision making. *Organizational Behavior and Human Decision Processes* 43 (3): 301–335.

——— 2000. *Business Dynamics: Systems Thinking and Modeling for a Complex World.* Boston, MA: Irwin McGraw-Hill.

Sternberg, R. 1997. Cognitive conceptions of expertise. In *Expertise in Context*, edited by P. Feltovich, K. Ford, and R. Hoffman, 149–162. Menlo Park, CA: The MIT Press.

Sweeney, L. B., and J. D. Sterman. 2000. Bathtub dynamics: Initial results of a systems thinking inventory. *System Dynamics Review* 16: 249–286.

Sweller, J., J. Van Merrienboer, and F. Paas. 1998. Cognitive architecture and instructional design. *Educational Psychology Review* 10 (3): 251–296.

Tindall-Ford, S., P. Chandler, and J. Sweller. 1997. When two sensory models are better than one. *Journal of Experimental Psychology. Applied* 3: 257–287.

Vera-Muñoz, S. C., M. Shackell, and M. Buehner. 2007. Accountants' usage of causal business models in the presence of benchmark data: A note. *Contemporary Accounting Research* 24 (3): 1015–1036.

THE ACCOUNTING REVIEW
Vol. 86, No. 3
2011
pp. 945–974

American Accounting Association
DOI: 10.2308/accr.00000046

Changes over Time in the Revenue-Expense Relation: Accounting or Economics?

Dain C. Donelson
Ross Jennings
John McInnis
The University of Texas at Austin

ABSTRACT: Dichev and Tang (2008) document a dramatic decrease over the last 40 years in the contemporaneous correlation between revenue and expense, along with an associated increase in earnings volatility and a decline in earnings persistence, suggesting a decline in earnings quality. We document that these changes are primarily attributable to an increase in the incidence of large special items. We then examine the extent to which this increase in special items is due to either more frequent real economic events related to special item recognition or to the adoption of new accounting standards. Our evidence suggests that changes in the frequency of economic events associated with special items have played a more important and sustained role relative to the role played by adoption of individual accounting standards. Finally, we find that the changing incidence of these economic events is at least in part related to the well-documented increase in competition in the U.S. economy over the last four decades.

Keywords: *earnings persistence; earnings volatility; matching; special items.*

Data Availability: *Data are available from public sources as identified in the text.*

JEL Classifications: *M41.*

I. INTRODUCTION

Dichev and Tang (2008; hereafter, DT) provide important evidence on changes over time in the relation between contemporaneous revenue and expense. For the 1,000 largest U.S. firms from 1967 to 2003, they find a decrease in the correlation between current-period revenue and current-period expense and an increase in the correlation between current-period revenue and expense from the prior and subsequent periods. These findings suggest a decline in "matching," such that an increasing amount of expense is being recognized before and after the period in which it affects revenue. DT suggest these changes could arise from either changes in economic factors or changes in accounting standards, and provide indirect evidence that changes

We thank Bill Kinney, Laura Wang, Yong Yu, Mark Trombley (editor), two anonymous referees, and workshop participants at the Chinese University of Hong Kong and The University of Texas at Austin for their helpful comments.

Editor's note: Accepted by Mark Trombley.

Submitted: August 2009
Accepted: November 2010
Published Online: April 2011

in accounting standards are primarily responsible. In this study, we build upon DT by first identifying which expense line items are responsible for changes in the revenue-expense relation. We then examine whether the responsible line items have been more affected by changes in certain economic activity or by changes in specific accounting standards.

Understanding the factors leading to changes in the revenue-expense relation is important for researchers, financial statement users, and accounting standard-setters for two reasons. First, these changes have important implications for earnings quality. As modeled by DT and supported by their empirical results, reporting related revenue and expense in different accounting periods adds noise to the measurement of earnings. This noise increases earnings volatility and decreases earnings persistence, consistent with a decline in earnings quality (see, e.g., Dechow 1994; Francis et al. 2004; Dechow and Schrand 2004; DT). Thus, the results in DT suggest a prolonged decline in the quality of earnings.

Second, identifying the source of these changes in earnings properties is important because the two explanations we investigate have different policy implications. If accounting standards are responsible for changes in the revenue-expense relation, then the resulting decline in the "usefulness" of accounting information is in direct opposition to the objectives of the Financial Accounting Standards Board's (FASB) conceptual framework. If this is the case, it would be important to understand which standards are most responsible and consider corrective action. Alternatively, if changes in economic activity are responsible, then standard-setters should consider this possibility when they evaluate current and future standards.

To determine which expense items are responsible for changes in the revenue-expense relation over time, we disaggregate total expenses into: cost of goods sold; selling, general, and administrative expense; depreciation; taxes; other income/expenses; and special items. Using a simple decomposition framework, we show that the decline in the relation between revenue and current expense is attributable primarily to a single income statement line item—special items. This line item consists mostly of asset impairments, restructuring charges, and gains/losses from asset sales. Not surprisingly, current-period special items have a much lower level of association with current-period revenue than other expenses. In addition, the "weight" of special items as a component of total expense has increased with the incidence of special items over time, decreasing the relation between current revenue and total current expense. We find that the most important changes in revenue and expense patterns are substantially reduced (by 90 percent in many cases) by either (1) excluding firm-years that report large special items (those greater than 1 percent of total assets), (2) using operating expense after depreciation (which excludes special items) to define total expenses, or (3) adding back special items to net income.

Next, we examine whether the increasing incidence of special items is due to changes in economic activity or changes in specific accounting standards. First, we provide evidence on whether the increasing incidence of special items is related to changes in the frequency of specific economic events associated with recognition of special items. We identify such economic events using five variables: negative employee growth; merger and acquisition activity; discontinued operations; negative revenue growth; and operating losses. These five variables represent real economic events whose frequencies are not likely to be influenced by accounting standards or by bookkeeping practices. Rather, they likely capture economic circumstances directly or indirectly associated with asset impairments, restructuring activities, and asset sales. We combine these variables into an index and show that this index both (1) increases throughout our study period, and (2) explains significant cross-sectional variation in the incidence of special items at the firm level. This finding provides strong evidence that the increase in special items is related to changes in the incidence of underlying economic events.

Next, we search for accounting standards adopted during our study period that are related to specific types of special items and find relatively few standards of this type. Among those that we

find, many appear to be either (1) reactive to changes in economic activity and accounting practice rather than leading such changes, or (2) too late in the study period to explain increases in special items prior to their adoption. In addition, we find that total special items increased in periods when no new accounting standards were adopted. Moreover, targeted tests of specific special items provide little evidence that adoption of specific accounting standards affected the frequency of these special items. Only for asset impairments do we find that implementation of a standard was associated with an increase in special items after controlling for a general time trend and our index of economic activity.

We also compare the effects of the specific accounting standards and our index of economic events on the increase in total special items. We find that changes in economic events play a more important role in the increasing incidence of special items over time relative to the role played by the adoption of specific accounting standards.

Finally, due to the apparently important role of economic events, we examine whether increases in our index of economic events are related to increasing competitive pressures in the economy documented by prior studies. We identify two measures of competitive pressure: industry concentration ratio and average probability of bankruptcy for the industry. We combine these measures into an index of competitive pressure and find that this index is (1) increasing over time throughout our study period, and (2) positively associated in the cross-section with our index of economic events related to special item recognition.

This study makes several important contributions. First, the results indicate that the deterioration in the contemporaneous relation between revenue and expense is due primarily to a dramatic increase in the frequency of reported special items, not to systematic issues across multiple line items in the income statement. Second, we provide strong evidence that an increase in the frequency of economic events associated with special item recognition plays an important role in the increasing frequency of special items. Third, the results also indicate that changes in specific accounting standards related to special items play, at best, a secondary role in the increasing frequency of special items. Thus, a primary implication of our results for standard-setters is that they should be aware of the relation between changes in the economic environment and the increasing incidence of special items over time.

II. BENCHMARK CHANGES IN THE REVENUE-EXPENSE RELATION

In this section, we provide benchmark evidence consistent with DT that the correlation between current-period revenue and current-period expense has declined over time. We begin by selecting a sample that is generally consistent with the sample in DT. First, we identify the 1,000 largest firms each year from 1967 to 2005 on the basis of total assets. Like DT, we select relatively large companies to confine our focus to changes in earnings properties among economically significant firms. This selection also helps maintain comparability throughout the sample period, as results are unaffected by the proliferation of smaller public firms later in the sample period. We require firms to have data available to calculate average earnings over successive, non-overlapping, two-year intervals ("two-year earnings") and the volatility of two-year earnings over the preceding ten years. We also require firms to have non-missing values for cost of goods sold, SG&A expense, income tax expense, and operating income after depreciation. We exclude financial firms, due to the difficulty in interpreting conventional income statement components for these firms.[1] Finally, we eliminate observations in the extreme 1 percent of the distribution each year of

[1] When we include financial firms in our sample, the results for the main analyses below are inferentially identical.

two-year earnings, two-year revenue and expense, two-year earnings volatility, current revenue, current expense, last-year expense, and next-year expense.[2] The sample has 32,645 firm-year observations between 1967 and 2005.

Next, we estimate the following cross-sectional regression by year, which is identical to the regression reported in DT (2008, Table 3):

$$Rev_{i,t} = \gamma_0 + \gamma_1 Exp_{i,t-1} + \gamma_2 Exp_{i,t} + \gamma_3 Exp_{i,t+1} + \varepsilon_{i,t}. \tag{1}$$

$Rev_{i,t}$ is total revenue for firm i for year t, and $Exp_{i,t}$ is total expense, computed as revenue minus net income before extraordinary items for firm i in year t, divided by average total assets for year t. These, and all other variables used in the study, are defined in the Appendix. The coefficient on current-period expense estimates the extent to which current-period expense varies with revenue in the current period, and the coefficient estimates on expense in the prior period and the next period estimate the extent to which expenses are scattered to adjacent periods.

Average coefficients are reported in Table 1, Panel A, for the first half (1967 to 1985) and second half (1986 to 2005) of the study period. The average coefficient estimate on contempora-

TABLE 1

Properties of Earnings-Related Variables over Time

Panel A: Relation of Revenue to Lagged, Current and Future Expense

	Exp_{t-1}	Exp_t	Exp_{t+1}
1967–1985	0.002	1.032	−0.030
1986–2005	0.089	0.895	0.025
Difference	0.087***	−0.137***	0.055***
t-statistic	(5.98)	(−5.54)	(4.04)

Panel B: Volatility and Persistence of Earnings

	Volatility	Persistence
1967–1985	0.022	0.805
1986–2005	0.036	0.680
Difference	0.014***	−0.125***
t-statistic	(9.77)	(−6.32)

*, **, *** Represent significance at the 10 percent, 5 percent, and 1 percent levels, respectively, in two-tailed tests.

For difference calculations only, t-statistics are in parentheses below the coefficient estimates.

This table presents properties of earnings-related variables between two time periods, 1967–1985 and 1986–2005. Panel A presents average annual coefficients from estimating the following model each in both time periods:

Model 1: $Rev_{i,t} = \gamma_0 + \gamma_1 Exp_{i,t-1} + \gamma_2 Exp_{i,t} + \gamma_3 Exp_{i,t+1} + \varepsilon_{i,t}$.

Panel B presents average firm specific volatility of *Earn* and average coefficients (persistence) from estimating the following model each year in both time periods:

Model 2: $Earn_{i,t} = \gamma_0 + \gamma_I Earn_{i,t-1} + \varepsilon_{i,t}$.

Variables are defined in the Appendix.

[2] These eliminations follow the sample selection procedures in DT. When we do not eliminate these observations, the decrease in the contemporaneous correlation between revenue and expense is somewhat more pronounced.

neous expense is 1.032 (0.895) in the early (late) period, and the difference of 0.137 is highly significant ($p < 0.01$). We also report increases in the lag (lead) coefficient from 0.002 (−0.030) in the early period to 0.089 (0.025) in the late period.

As described by DT, the decline in the contemporaneous association of revenue and expense is likely to have implications for the volatility and persistence of earnings, which in turn have important implications for the quality of earnings. Specifically, scattering expenses related to current revenue to other time periods adds "noise" to reported expenses, resulting in more volatile and less persistent earnings. To see this, we follow DT and compute the firm-specific volatility of earnings (as measured by its standard deviation) over rolling five-year windows, and then average this measure across firms each year and then across years within each sub-period. We also follow DT and estimate earnings persistence using the following regression:

$$Earn_{i,t} = \gamma_0 + \gamma_1 Earn_{i,t-1} + \varepsilon_{i,t}. \tag{2}$$

The results for both earnings volatility and earnings persistence are presented in Panel B of Table 1.[3] The volatility of earnings increases from 0.022 in the early period to 0.036 in the later period, and the change is significant ($p < 0.01$). The earnings persistence coefficient (γ_1 from model 2) decreases from 0.805 in the early period to 0.680 in the later period, and this difference is also significant ($p < 0.01$). These results indicate that the decline in the contemporaneous relation between revenue and expense documented in Panel A of Table 1 has produced an increase in the volatility of earnings and a decrease in the persistence of earnings from the early years to the later years in our study period, consistent with a decline in earnings quality (see, e.g., Dechow 1994; Francis et al. 2004; Dechow and Schrand 2004).

Overall, the results reported in Table 1 are virtually identical to those reported by DT (2008, Tables 3–5). In the next section we examine which income statement line items are responsible for the decline in the revenue-expense relation documented in Table 1.

III. DISAGGREGATED EXPENSES AND THE REVENUE-EXPENSE RELATION

We begin our analysis with a decomposition of the change in average coefficients from Equation (1) between the early period and the late period. To do this, we divide total expense in Equation (1) into six components: cost of goods sold (*COGS*), selling, general, and administrative expense (*SGA*), depreciation expense (*DEPR*), tax expense (*TAX*), other expenses (*OTH*), and special items (*SI*). *SGA* includes R&D and advertising expense. Estimating Equation (1) implicitly constrains all of the coefficients on these six components that comprise total expense for a given year to be identical. Alternatively, the following regression allows the coefficients on these six components for the current year to vary freely:

$$Rev_{i,t} = \gamma_0 + \gamma_1 Exp_{i,t-1} + \gamma_{2a} COGS_{i,t} + \gamma_{2b} SGA_{i,t} + \gamma_{2c} DEPR_{i,t} + \gamma_{2d} TAX_{i,t} + \gamma_{2e} OTH_{i,t} + \gamma_{2f} SI_{i,t} + \gamma_3 Exp_{i,t+1} + \varepsilon_{i,t} \tag{3}$$

A decomposition similar to that in Kee (2009) demonstrates that the annual coefficient estimate for Exp_t (γ_2) in Equation (1) is a weighted average of the parameters on the individual component estimates (γ_{2a} through γ_{2f}) from Equation (3).[4] Thus, the decline we observe in the coefficient estimate on aggregate expense may occur for one or both of two main reasons. First, it

3 To mitigate the influence of extreme observations in our volatility and persistence tests, we eliminate observations in the extreme 1 percent of the distribution each year of earnings volatility and current- and prior-year earnings.

4 The decomposition in Kee (2009) is limited to two components that are assumed to be positively correlated. Our analysis is based on six components whose pair-wise correlations may be either positive or negative. A proof that the analysis in Kee (2009) extends to our case without loss of generality is available from the authors on request.

is possible that the coefficient estimates for some (or all) of the individual expense components decline between the two periods. Second, it is also possible that the "weights" of the individual components have shifted over time, away from components with larger coefficient estimates and toward components with smaller coefficient estimates. As demonstrated by Kee (2009), the weights are related to the importance of a particular component—specifically, how much of the variance of total expense is attributable to each expense component—and must sum to 1 by construction.[5]

Table 2 summarizes the changing coefficients and weights of the expense components. The first set of rows provides the average coefficient estimates for the expense components for the early and late periods as well as the change between periods. The next set of rows provides the average weight for each component for each period as well as the change between periods. It is clear from Table 2 that special items has the lowest average coefficient estimate in both periods and that its average coefficient estimate has declined the most from the early period (0.483) to the late period (0.197). At the same time, we observe a dramatic shift in weight toward special items (from 0.009 to 0.129) and away from high-coefficient components like *COGS* and *TAX*. The increased weight on special items is consistent with prior studies that document an increase over time in one-time, special charges such as asset write-offs, asset sales, and restructuring charges (Elliott and Hanna 1996; Collins et al. 1997; Fairfield et al. 2009).[6]

TABLE 2

Decomposition Analysis
Current Year Expense

		COGS	*SGA*	*DEPR*	*TAX*	*SI*	*OTH*
Coefficients	1967–1985	1.004	1.017	1.058	1.891	0.483	0.584
	1986–2005	0.996	1.030	0.853	2.123	0.197	0.480
	Difference	−0.009	0.014	−0.205	0.233	−0.286	−0.105
Weights	1967–1985	0.834	0.106	0.006	0.039	0.009	0.007
	1986–2005	0.711	0.122	0.013	0.011	0.129	0.013
	Difference	−0.122	0.016	0.008	−0.027	0.120	0.006

We decompose the aggregate coefficient estimates on current expense from Panel A of Table 1 into six income statement components and their respective weights, using the following model:

$$\text{Model 3: } Rev_{i,t} = \gamma_0 + \gamma_1 Exp_{i,t-1} + \gamma_{2a}COGS_{i,t} + \gamma_{2b}SGA_{i,t} + \gamma_{2c}DEPR_{i,t} + \gamma_{2d}TAX_{i,t} + \gamma_{2e}OTH_{i,t} + \gamma_{2f}SI_{i,t} + \gamma_3 Exp_{i,t+1} + \varepsilon_{i,t}$$

Kee (2009) demonstrates that the coefficient on current expense in Panel A of Table 1 (γ_2) is simply a weighted average of coefficients on the expense components (γ_{2a} through γ_{2f}) in (3). The weight for any component$_i$ equals Cov(*Exp*, *component*$_i$)/Var(*Exp*), where *Exp* is total current expense (Kee 2009).

[5] The weight for any component$_i$ equals Cov(*Exp*, *component*$_i$) / Var(*Exp*), where *Exp* is total current expense (see Kee [2009] for a proof). This is the coefficient one obtains from an OLS regression of *component*$_i$ on total expense. Since Var(*Exp*) = Σ_i Cov(*Exp*, *component*$_i$), the weights must sum to 1.

[6] A recent study by Cready et al. (2010) documents that the persistence of special items varies positively with the frequency of reported special items in the prior three years, indicating that for some firms these nonrecurring items are more recurring than for other firms. However, their evidence indicates that even for firms with five or more special items

These results are consistent with the decline we observe for the coefficient estimate on contemporaneous expense in Equation (1)—the low (and decreasing) coefficient estimate on special items increases in importance (weight) and drags down the coefficient on total expense. In contrast, the results for the other expense components are not consistent with the decline in the coefficient estimate on contemporaneous expense in Equation (1). For the other expense components, either the change in the estimated coefficient from the early period to the late period is very small or increasing (*COGS* and *SGA*) or the weight of the component is low in both periods (*DEPR*, *TAX*, and *OTH*).[7]

We perform four analyses to corroborate the critical role of special items in changes in the revenue-expense relation and the volatility and persistence of earnings reported in Table 1. First, we repeat all analyses in Table 1 after deleting observations with a large special item that would affect any of the results, where "large" is defined as 1 percent of average total assets for the year.[8] The results, which are reported in Table 3, Panel A, demonstrate the substantial effect large special items have on the earnings properties. After excluding observations with large special items, the coefficient estimate for current-period expense from Equation (1) is not significantly different between the early and late periods ($p = 0.78$), as compared to a significant decline of 0.137 in Table 1 ($p < 0.01$). In addition, between the early and late periods volatility (persistence) increases (decreases) by only 0.002 (0.003), as compared to 0.014 (0.125) in Table 1. Although the increase in volatility remains significant at the 0.01 level after excluding observations with large special items, the magnitude is only 14 percent of the magnitude reported in Table 1.

As a second method of evaluating the effect of large special items, we sort observations in the later period into those with and without large special items. There are not enough observations with large special items during the first years of the early period to conduct this analysis in the early period. The results, which are reported in Table 3, Panel B, demonstrate the dramatic effect of special items. For observations without special items, the late period coefficient estimate for contemporaneous expense from Equation (1) is 1.046, which is similar to the early period coefficient estimate reported in Table 1 (1.032). In contrast, the late period contemporaneous expense coefficient for observations with large special items is 0.752, which is significantly less than the estimate for observations without large special items ($t = 8.44$; $p < 0.01$). In addition, observations with large special items have significantly larger (smaller) estimates for volatility (persistence), demonstrating the effect of special items on these earnings properties.

---

in the previous twelve quarters, the persistence of those special items is still substantially below that of earnings adjusted for special items (see Cready et al. 2010, Table 3). Thus, while the recurring nature of some special items might partially mitigate the effects we report in Table 2, such mitigation is likely to be slight.

7 Although the weight on *COGS* declines substantially from the early to the late period in Table 2, we view the decline in the revenue-expense relation as a function more of special items rather than *COGS*. The weights must sum to 1, such that a dramatic increase in relative importance of special items must mechanically be offset by a decline in relative importance of other components. This issue is analogous to the product mix problem in management accounting where a company has a low-margin product (LM) and a high-margin product (HM). Assume that HM volume grows 10 percent and LM volume doubles, shifting the mix away from HM and toward LM and reducing overall profit margin. In this case it is logical to attribute the decline in profit margin to LM, the product whose growth was extreme. Similarly, in our case, we attribute the weight effect to special items, the component whose growth was extreme. In an untabulated analysis we find that more than 90 percent of the change in the coefficient estimate on aggregate expense in Equation (1) is due to the combined effects of the change in the coefficient estimate and the change in the weight of special items (details available from the authors on request).

8 Specifically, we re-estimate Equation (1) after deleting all observations with a large special item in the current year (5,330 observations), we re-estimate the volatility measures after deleting observations with a large special item in any of the prior five years (12,145 observations), and we re-estimate the persistence regression after deleting observations with large special items in the current or previous years (7,728 observations). In addition, we repeat these analyses with a smaller, common sample formed by the intersection of these three subsamples, and results are very similar. Our definition of "large" is consistent with that used by Elliott and Hanna (1996).

TABLE 3

Properties of Earnings-Related Variables over Time Adjusting for Large Special Items

Panel A: Early versus Late Period, Excluding Observations with Large Special Items

	Exp_{t-1}	Exp_t	Exp_{t+1}
1967–1985 (No Large *SI*)	−0.003	1.042	−0.036
1986–2005 (No Large *SI*)	0.010	1.046	−0.046
Difference	0.013*	0.004	−0.010*
t-statistic	(1.97)	(0.42)	(−1.73)
	Volatility	**Persistence**	
1967–1985 (No Large *SI*)	0.020	0.822	
1986–2005 (No Large *SI*)	0.022	0.819	
Difference	0.002***	−0.003	
t-statistic	(3.11)	(−0.15)	

Panel B: Late Period, With versus Without Large Special Items

	Exp_{t-1}	Exp_t	Exp_{t+1}
1986–2005 (No Large *SI*)	0.010	1.046	−0.046
1986–2005 (Large *SI*)	0.165	0.752	0.076
Difference	0.155***	−0.294***	0.123***
t-statistic	(8.10)	(−8.44)	(5.53)
	Volatility	**Persistence**	
1986–2005 (No Large *SI*)	0.022	0.819	
1986–2005 (Large *SI*)	0.032	0.529	
Difference	0.010***	−0.290***	
t-statistic	(15.94)	(−11.43)	

Panel C: Early versus Late Period, Based on Operating Income

	Exp_{t-1}	Exp_t	Exp_{t+1}
1967–1985	−0.012	1.067	−0.046
1986–2005	0.011	1.059	−0.058
Difference	0.024**	−0.008	−0.012
t-statistic	(2.85)	(−0.63)	(−1.35)
	Volatility	**Persistence**	
1967–1985	0.035	0.819	
1986–2005	0.034	0.813	
Difference	−0.001	−0.006	
t-statistic	(−0.50)	(−0.15)	

(continued on next page)

Panel D: Early versus Late, Adding Back Special Items

	Exp_{t-1}	Exp_t	Exp_{t+1}
1967–1985	−0.003	1.038	−0.031
1986–2005	0.023	1.024	−0.038
Difference	0.025***	−0.013*	−0.007
t-statistic	(3.99)	(−1.70)	(−1.09)

	Volatility	Persistence
1967–1985	0.021	0.822
1986–2005	0.027	0.786
Difference	0.006***	−0.036*
t-statistic	(6.56)	(−1.87)

*, **, *** Represent significance at the 10 percent, 5 percent, and 1 percent levels, respectively, in two-tailed tests.

For difference calculations only, t-statistics are in parentheses below the coefficient estimates.

This table presents properties of earnings-related variables for two time periods, 1967–1985 and 1986–2005, utilizing four different methods to control for the presence of special items. All regression specifications are equivalent to those estimated in Table 1 except as noted below.

In each panel, the first set of results presents average annual coefficients from estimating:

Model 1: $Rev_{i,t} = \gamma_0 + \gamma_1 Exp_{i,t-1} + \gamma_2 Exp_{i,t} + \gamma_3 Exp_{i,t+1} + \varepsilon_{i,t}$.

In each panel, the second set of results presents average firm-specific volatility of *Earn* and average coefficients (persistence) from estimating:

Model 2: $Earn_{i,t} = \gamma_0 + \gamma_1 Earn_{i,t-1} + \varepsilon_{i,t}$.

Panel A presents results for the early and late time periods after excluding firm-years with large special items (defined as greater than 1 percent of average total assets).

Panel B presents results for the late time period, comparing observations with large special items to observations without large special items.

Panel C presents results for the early and late time periods based on operating income after depreciation as the earnings measure.

Panel D presents results for the early and late time periods after adding back special items to earnings before discontinued operations.

Variables are defined in the Appendix.

Our third and fourth analyses avoid the deletion of observations and use alternative earnings measures to evaluate the effect of special items. In the third analysis, we replace net income before extraordinary items with operating income after depreciation (but before special items) and we replace total expense with revenue minus operating income after depreciation. In the fourth analysis, we add special items back to net income before extraordinary items and replace total expense with revenue minus this modified income figure.

The results for the third (fourth) analysis are reported in Panel C (D) of Table 3. In Panel C (D), the change in the coefficient estimate for current-period expense from Equation (1) from the early to the late period is −0.008 (−0.013), as compared to −0.137 in Table 1. The difference in Panel C is 6 percent of the original difference and is not significantly different from 0 (p = 0.52), while the difference in Panel D is 10 percent of the original difference and is weakly significant (p = 0.10). Also, in Panel C, the changes in both volatility and persistence from the early to the late period are small when compared to the results in Table 1, and neither change is statistically significant (p = 0.61 and 0.59, respectively). In Panel D, when special items are added back to

earnings, the increase (decrease) in volatility (persistence) is 43 percent (29 percent) of the increase (decrease) reported in Table 1, and both changes are significant ($p < 0.01$; $p = 0.07$).[9]

We also observe that adjusting for special items in the current period affects the coefficient estimates on aggregate expense in the surrounding periods. Specifically, the increases in the coefficient estimates on non-contemporaneous expenses are attenuated or reversed. In Table 1, these coefficient estimates likely increased because the increased incidence of special items added noise to the contemporaneous relation between revenue and total current expense over time. This noise was likely compensated for by an increase in the relation between current revenue and non-current expense because (1) these expenses are highly correlated with prior period and next period revenue, and (2) revenue is highly persistent from one period to the next. Thus, all of the patterns we observed in Table 1 are substantially reduced by each of our adjustments for special items in Table 3.

Taken together, these results suggest that the significant patterns documented by DT and reported in Table 1 are due primarily to the increase in the weight of special items as a component of total expense. In the next sections, we provide further evidence on the increase in special items over time and explore explanations for this phenomenon.

IV. CHANGES OVER TIME IN THE INCIDENCE OF SPECIAL ITEMS

To provide a general view of the extent to which special items are increasing during the sample period, we first plot the percent of sample observations that reported a large special item for each year from 1967 to 2005 in Figure 1. The incidence of large special items is below 5 percent before 1973 and rises to almost 50 percent in 2001. This pattern is consistent with Elliott and Hanna (1996, Figure 1) and Fairfield et al. (2009, Figure 1).

Next, we provide additional descriptive evidence on changes in the frequency of commonly reported special items using data reported by *Accounting Trends and Techniques*, an annual publication of the American Institute of Certified Public Accountants. Each volume contains examples of corporate disclosures and tabulations of reporting practices for 600 industrial, retail, and service companies chosen from the Fortune 1000.[10] Throughout most of our study period, *Accounting Trends and Techniques* has included a tabulation of various types of other revenue and gains and other expenses and losses disclosed by firms. We summarize these tabulations in Table 4, Panel A for gains and in Panel B for losses. We tabulate only items that the Compustat manual includes in its definition of special items. Each cell reports the percent of the 600 firms surveyed by *Accounting Trends and Techniques* reporting that item that year. The columns sum to more than 100 percent because a company may have more than one type of gain or loss (e.g., sale of assets and litigation settlement), and a gain or loss may count in more than one category (e.g., write-down of assets and restructuring of operations).[11]

The results in Table 4, Panel A, indicate a steady increase in gain special items between 1975 and 2005, with average total incidence for the first (last) three years of 9 percent (54 percent),

---

[9] The tests reported in Panel D of Table 3 are similar to a robustness test conducted by DT. DT report that the original change in the coefficient estimate on contemporaneous expense from Equation (1) of -0.149 falls to -0.081 when special items are added back to income. We find that our original change of -0.137 falls to -0.013 after adding back special items, which is consistent with our decomposition results in Table 2 and our results from the additional adjustments for special items in Table 3.

[10] There is a substantial overlap between this group and our sample. For example, for 2002, of the 750 firms in our sample, 350 (47 percent) were also included in the firms surveyed by *Accounting Trends and Techniques*.

[11] Compustat and *Accounting Trends and Techniques* use similar procedures to identify special items. In both cases, items are identified as a special item when they are disclosed either in the main financial statements or in the financial statement notes. See Burgstahler et al. (2002) for discussion of Compustat procedures.

FIGURE 1
Large Special Items over Time

This figure graphs the incidence of large special items (those greater than 1 percent of average total assets) per year in the sample.

which is about a sixfold increase. Among the individual gain special items, only asset sales have a consistently high frequency, rising from 8 percent in 1975 to 37 percent in 2005.

The incidence of loss special items (Panel B) is rising over time more rapidly than the incidence of gain special items. The most common loss special items are asset write-downs, restructuring charges, and losses on asset sales. Across all loss special items, the average total incidence for the first (last) three years reported in the table is 5 percent (126 percent), which is more than a 25-fold increase. Even if we eliminate write-downs and asset sales as being included in restructurings in later years, loss special items still increase 16-fold. Loss special items are increasing faster than gain special items both because the types of items being reported are increasing and because of increases in the frequency of particular items being reported. For example, only two types of loss special items are reported in *Accounting Trends and Techniques* for 1980, while for 2005 eight types of loss special items have reported amounts. Also, in 1980 only 3 percent of firms report an asset write-down, but this percentage increases to 35 percent in 2005.

TABLE 4

Losses and Gains from *Accounting Trends and Techniques*

Panel A: Gains

Special Items (%)	05	04	03	02	01	00	99	98	97	96	95	94	93	92	91	90
Asset sale	37	33	33	32	30	16	26	25	27	21	23	20	17	20	18	22
Liab reversal	12	12	11	11	8	6	5	4	2	2	3	nc	nc	nc	nc	nc
Litigation	5	6	6	3	2	3	1	3	2	1	2	2	3	3	2	2
Emp benefit	1	*	nc	nc	nc	nc	nc	nc	nc	nc	nc	*	1	1	2	5
Debt ext	1	2	2	1	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	1
Total	56	53	52	46	40	25	32	32	31	24	28	22	21	24	22	29

Panel A: Gains, continued

Special Items (%)	89	88	87	86	85	84	83	82	81	80	79	78	77	76	75
Asset sale	27	24	25	26	25	20	22	20	20	16	15	12	11	8	8
Liab reversal	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc
Litigation	2	2	2	1	2	2	1	nc	nc	nc	nc	nc	nc	nc	nc
Emp benefit	3	4	5	5	1	1	nc	nc	nc	nc	nc	nc	nc	nc	nc
Debt ext	1	1	1	1	3	5	4	10	nc	nc	nc	nc	nc	nc	nc
Total	33	31	33	33	31	28	27	30	20	16	15	12	11	8	8

Panel B: Losses

Special Items (%)	05	04	03	02	01	00	99	98	97	96	95	94	93	92	91	90
Write-down	35	32	34	36	39	27	25	29	18	19	19	11	11	8	11	12
Asset sale	12	12	13	16	12	7	8	5	7	5	4	4	5	6	7	6
Restructuring	42	38	37	44	48	26	27	33	24	23	22	17	32	23	24	19
Litigation	10	10	8	7	6	4	4	6	4	6	4	5	4	3	3	3
Cleanup	6	6	4	4	3	2	3	5	5	4	5	5	5	4	5	5
Purch R&D	3	3	2	3	4	3	3	7	3	nc	nc	nc	nc	nc	nc	nc
Int impair	12	10	12	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc

(continued on next page)

Panel B: Losses Special Items (%)	05	04	03	02	01	00	99	98	97	96	95	94	93	92	91	90
Debt ext	13	14	10	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc
Total	133	125	120	110	112	69	70	85	61	57	54	42	57	44	58	45

Panel B: Losses, continued Special Items (%)	89	88	87	86	85	84	83	82	81	80	79	78	77	76	75
Write-down	8	7	2	10	8	7	6	6	5	3	3	2	3	3	0
Asset sale	5	4	5	5	5	4	5	4	3	4	4	5	5	4	0
Restructuring	16	12	14	22	17	12	13	15	9	8	nc	nc	nc	nc	nc
Litigation	2	3	2	2	3	2	2	nc	nc	nc	nc	nc	nc	nc	nc
Cleanup	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc
Purch R&D	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc
Int impair	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc
Debt ext	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc	nc
Total	31	26	23	39	33	25	26	25	17	15	7	7	8	7	0

Panel C: Abbreviations Used for Line Items from the Tables in *Accounting Trends and Techniques*

Gains
Special Items
- Asset sale: Sale of assets
- Liab reversal: Liability accrual reduced
- Litigation: Litigation settlements
- Emp benefit: Employee benefit/pension related
- Debt ext: Debt extinguishment

Not Tabulated:
Non-Special Items (per Compustat manual)
- Interest
- Equity in earnings of investees

Losses
Special Items
- Write-down: Write-down of assets
- Asset sale: Sale of assets
- Restructuring: Restructuring of operations
- Litigation: Litigation settlements
- Cleanup: Environmental cleanup
- PRD: Purchased R&D
- Intan Impair: Impairment of intangibles
- Debt ext: Debt extinguishment

Not Tabulated:
Non-Special Items (per Compustat manual)
- Foreign currency transactions
- Equity in losses of investees

(continued on next page)

Panel C: Abbreviations Used for Line Items from the Tables in *Accounting Trends and Techniques*

Dividends	Minority interests
Foreign currency transactions	Intangible asset amortization
Change in fair value of derivatives	Sale of receivables
Royalties	Royalties
Rentals	Merger costs: Merger costs
Insurance recoveries	Distributions of preferred securities of subsidiary trust
Public offering of subsidiary stock	Change in fair value of derivatives
Revenue of nonhomogeneous operations	Start-up costs
Sale of tax benefits	

"nc" Indicates this data were not compiled by *Accounting Trends and Techniques* and an asterisk (*) indicates that this item is less than 0.5 percent.

Panels A and B: Each cell reports the percent of firms out of the 600 surveyed by *Accounting Trends and Techniques* that reported the item in that row. Identification of Compustat special items is based on items specifically identified in the Compustat manual. However, additional items may be included if the firm identifies them as special items, so this list is a guide only, not determinative (see Standard & Poor's 2003).

V. ALTERNATIVE EXPLANATIONS FOR THE INCREASING FREQUENCY OF SPECIAL ITEMS

In general there are three potential reasons for the increase in special items over the past several decades. First, changes in economic activity could result in firms recording more special items than before. For example, it could be more common in the 2000s than in the 1960s for firms to experience impaired assets, sell an asset before it is used up, or lay off employees. Such changes in economic activity can lead to an increase in special items even when there has been no change in formal accounting standards. Second, new accounting standards could either require firms to record a special item that was previously ignored or provide more specific guidance that effectively encourages firms to record a special item more frequently than before. Below, we discuss specific accounting standards that relate to special items. Third, changes in accounting and disclosure practice can lead to firms recording more special items than previously even when economic events and formal accounting standards are not changing. For example, firms may begin to more actively write down assets that have been impaired rather than continuing to depreciate them according to the original depreciation schedule.[12] Such changes in accounting practice could occur if the financial reporting incentives for the firm or manager change. However, these changes are difficult to disentangle from changes in accounting practice that are precipitated by changes in economic activity or formal accounting standards.[13]

Despite a significant amount of research on special items, there is little evidence on which of these three potential reasons for the increase in special items is most important. Indeed, Frankel (2009, 239) states that "we do not understand why the use of special items has increased." In this section, we provide evidence on these potential reasons for the increase in special items in recent decades. Because changes in accounting practice are difficult to separate from the other two reasons, we focus on economic activity and accounting standards as two alternative, non-exclusive explanations for the increase in special items and the resulting effect on the properties of accounting earnings.

DT also consider accounting standards and economic explanations for their findings by testing whether their main results obtain for both accrual and cash flow expense measures. They find that the changes in the properties of earnings are attributable primarily to accruals and conclude that "accounting-based factors are a substantial and perhaps even primary determinant of the documented temporal patterns of earnings characteristics, while real-economy changes play a secondary role" (DT 2008, 1453).[14] Our finding that special items are responsible for the change in earnings properties is consistent with DT's finding with respect to accruals because special items are almost exclusively accruals (Dechow and Ge 2006). However, it is not clear *ex ante* whether more frequent accruals for restructuring charges, impairments, and asset sales are driven predominantly by new accounting standards or by economic factors.

---

[12] Alternatively, firms may begin to more transparently disclose write-downs that they were already recording but not separately disclosing. We note, however, that more transparent disclosure of a constant frequency and magnitude of special items would not explain the changes from the early to the late period reported in Table 1.

[13] Recent research by McVay (2006) and Fan et al. (2010) provides evidence that managers opportunistically manage recurring firm performance by including recurring charges as one-time special items. However, neither of these studies provides evidence on whether such classification shifting is either increasing over time or affects the frequency of special items. Moreover, such classification shifting, which results in special items that are more persistent, is likely to partially mitigate the disruptive effect of special items on the revenue-expense relation documented in Tables 2 and 3. Thus, while classification shifting is an important issue in accounting for special items, it is not likely to contribute to the role of special items in the revenue-expense relation.

[14] DT also examine whether changes in industry composition in their sample drive their results as another potential economic explanation. We find that special items are increasing across virtually all industries (untabulated).

The Role of Economic Events

In this subsection we examine whether an increase in economic activity associated with special items has contributed to the increasing frequency of special items. To examine the role of changes in economic activity, we construct an index that includes specific economic events that cannot arise from bookkeeping practices alone, but are frequently associated with firms recording special items. We then examine whether this index is associated with the frequency of special items across firms and over time.

We begin by identifying five firm-year indicator variables that represent real economic events that are likely to be associated with the most common special items in Table 4. Our first measure (*Emp*) takes the value of 1 if the firm has experienced negative employee growth that year, and 0 otherwise. Negative employee growth is often the result of contracting business, and can be the direct result of restructuring layoffs that result in a special item. Our second and third measures are merger and acquisition activity (*M&A*) and discontinued operations (*Disc*). These variables reflect underlying growth or contraction through firm entry and exit in markets. Firms that enter or exit new markets or product lines are also likely to restructure operations or sell or write off assets, so we expect *M&A* and *Disc* to be related to the incidence of special items. The fourth and fifth measures indicate if the firm has an operating loss (*Loss*) or if the firm has negative sales growth (*Grow*). Firms that experience an operating loss or declining sales often write off assets or restructure their businesses to become more competitive, so these variables are also expected to be associated with the probability of recording a special item.

To provide evidence that these economic event variables are linked to the incidence of special items, we combine them into a single measure and examine the relationship of this measure to the incidence of special items. We give each observation a point if the observation has a 1 for each indicator variable, and then sum the points across the five economic event variables for each firm-year observation and label that sum *E-Score*. Thus, firm-year observations with an *E-Score* of 0 (5) have a 0 (1) for each indicator variable, indicating a low (high) level of economic activity associated with special items.

To successfully capture economic events associated with the increasing incidence of special items, this index should increase over time and should also be associated with the cross-sectional probability of recording a special item. To examine whether the index is increasing over time, we calculate the average *E-Score* each year during the study period and then average the annual averages for four sub-periods, 1967–1974, 1975–1984, 1985–1996, and 1997–2005. The results are reported in the first row of Table 5, Panel A. The average *E-Score* increases steadily from 0.73 for the first sub-period to 1.15 for the last sub-period, an increase of nearly 60 percent. The correlation between average *E-Score* and a time trend is 0.55 ($p < 0.01$).[15]

To provide further evidence that *E-Score* is increasing over time, we divide the sample each year into three groups on the basis of *E-Score*. The "high" group consists of observations with *E-Score* equal to 3, 4, or 5, the "middle" group consists of observations with *E-Score* equal to 1 or 2, and the "low" group consists of observations with *E-Score* equal to 0. The next two rows in Panel A report that the middle (high) *E-Score* group has 50 (4) percent of the observations in the first sub-period, rising to 58 (11) percent of the observations in the last sub-period, confirming the increase in *E-Score* over time.

The results for each year are presented graphically in Figure 2, along with fitted trend lines, which indicate that the incidence of large special items for the high *E-Score* group is much higher than for the low group. Further, the incidence for the high group is growing at more than twice the rate for the low group. The trend line slope for the high group is 0.017 (i.e., an average increase

[15] We also observe statistically significant (at least $p < 0.10$) increases over time for all five components of *E-Score*.

TABLE 5
E-Score Classifications and Special Items

Panel A: *E-Score* over Time

	Mean 67–74	Mean 75–84	Mean 85–96	Mean 97–05	Time Trend	p-value
Level of *E-Score*	0.73	0.90	1.03	1.15	0.55	<0.01
Incidence of Middle *E-Score* Observations	0.50	0.52	0.54	0.58	0.49	<0.01
Incidence of High *E-Score* Observations	0.04	0.07	0.10	0.11	0.54	<0.01

Panel B: Cross-Sectional Relation between *E-Score* and Special Item Incidence

	Low *E-Score*	Middle *E-Score*	High *E-Score*	Diff
1967–1974	0.7	1.9	5.1	**4.4**
1975–1984	5.0	10.8	20.1	**15.1**
1985–1996	13.4	26.6	44.7	**31.2**
1997–2005	22.6	38.0	55.7	**33.2**
Diff	**21.9**	**36.1**	**50.7**	

We rank each firm-year observation on the incidence of five economic event variables from 1967 to 2005: negative employee growth, M&A, discontinued operations, negative revenue growth, operating losses. We give a firm-year observation one point for each indicator variable coded as 1. We then sum the points across the five variables for each firm-year observation, and label that sum *E-Score*. Thus, across observations, the *E-Score* can conceivably vary from 0 points (indicating the presence of no economic events) to 5 points (indicating the presence of all economic events). We then divide the sample each year into three groups: 0 points for the "Low" group (roughly bottom 25 percent of sample), 1 to 2 points for the "Middle" Group, and 3–5 points for the "High" group (roughly top 10 percent of sample).

In Panel A, we plot mean values of *E-Score* over four sub-periods between 1967–2005, along with a correlation with this time-series and a trend variable.

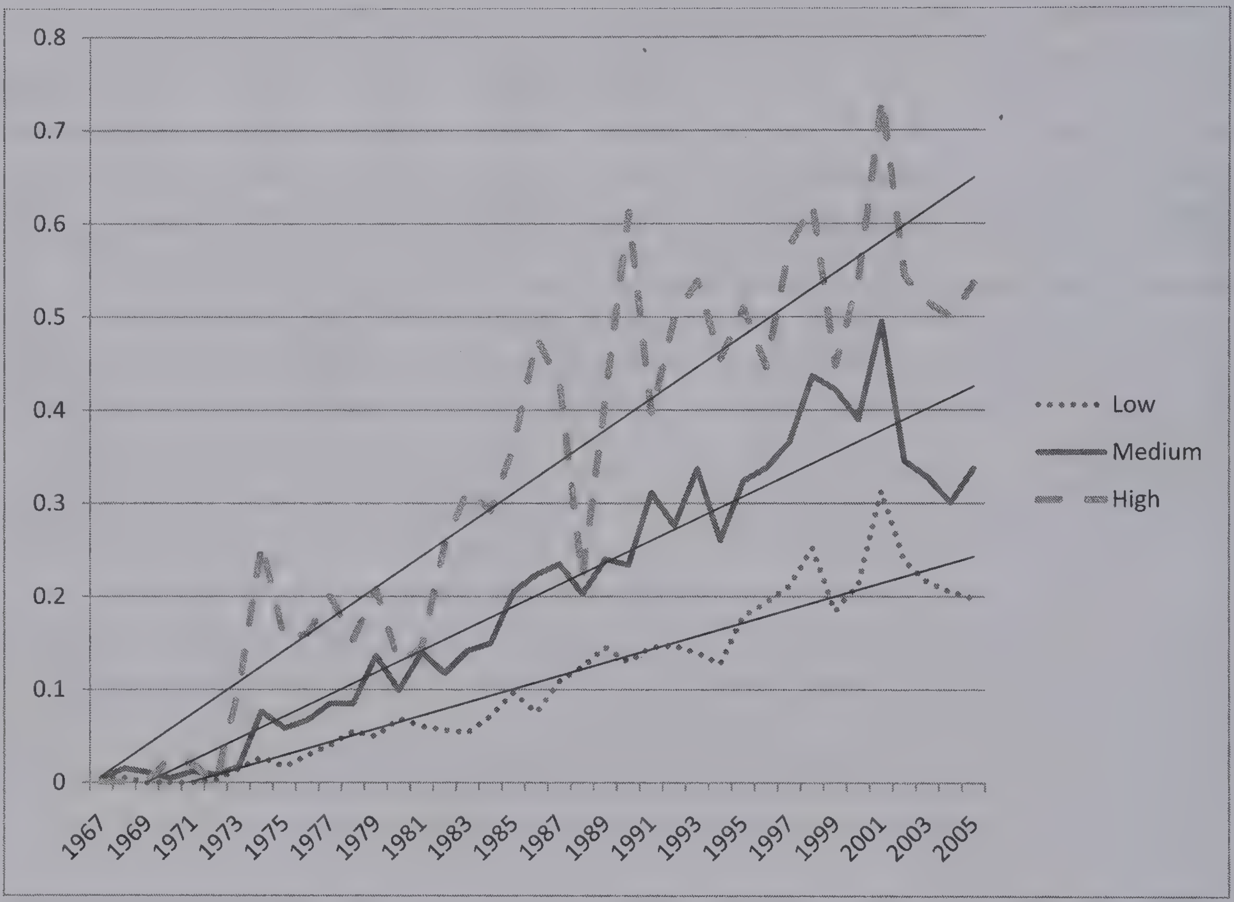
In Panel B, we compute the incidence of large special items for each of the three groups each year and then average over the four sub-periods. Bolded differences are significant at the 1 percent level (two-sided).

in the incidence of special items of 1.7 percent per year), compared with a slope of 0.007 for the low group, and the difference is significant at $p < 0.01$ (untabulated). This simple analysis provides strong evidence that our index of economic events is increasing over time and that it captures cross-sectional variation in the incidence of special items.

To provide further evidence that *E-Score* is increasing over time and is associated with the cross-sectional probability of recording a special item, we compute the average incidence of special items for the three *E-Score* groups for each of four sub-periods. The results, reported in Table 5, Panel B, reveal two clear patterns. First, the rows indicate that the average incidence of large special items for the high *E-Score* firms is more than double the incidence for the low *E-Score* firms in all four sub-periods. For example, in the most recent period, from 1997 to 2005, the incidence of large special items for firms in the high *E-Score* group is 55.7 percent, as compared to only 22.6 percent for the firms in the low *E-Score* group. These patterns indicate that *E-Score* is positively associated with the incidence of special items.[16]

[16] To document that variation in *E-Score* is directly related to deterioration in the revenue-expense relation, we also re-estimate Equation (1) each year separately for each of the three *E-Score* groups. In untabulated results we find that the coefficient estimate on contemporaneous expense declines overall as *E-Score* increases: 0.96, 0.98, and 0.84 for the low, middle and high *E-Score* groups, respectively.

FIGURE 2
Firm-Specific Economic Variables and Special Items



This figure graphs the incidence of special items for firms based on *E-Score* by year. *E-Score* is defined in Table 5. Firms in the "High" group have a score of 3 or higher, firms in the "Medium" group have a score of 1 or 2, while firms in the "Low" group have a score of 0. Scores are calculated by ranking each firm-year observation on each of the five economic variables in Table 5 from 1967 to 2005, and giving the observation one point for each indicator variable. We then sum the points across the five variables for each firm-year observation.

Second, the columns indicate that the incidence of special items increases through time for all three groups, and the differences between the incidence for the early sub-period and the late sub-period are significant at the 0.01 level for all three groups (last row). Thus, special items become more common even for relatively fixed levels of our economic activity index, indicating that *E-Score* does not capture all determinants of special items. This is reinforced by the variation around the fitted lines in Figure 2. One explanation for this finding is that changes in economic activity not captured by the five simple variables comprising *E-Score* are responsible for increases in special items. An alternative explanation is that, holding economic events related to special items constant, other factors, such as new accounting standards, led firms to record more special items. We examine this explanation in the next subsection.

The Role of Accounting Standards

To examine the role of accounting standards, we identify the specific standards that were adopted during our study period that are most likely to have had an effect on the frequency of

special items. We then examine the frequency of special items both in the years (1) prior to implementation of these standards (when they had no effect,) and (2) after implementation of these standards (when the effect, if any, should take place).

The general accounting standard that applied to special items for most of the study period is APB Opinion No. 30 (AICPA 1973) on reporting the results of operations, which became effective in 1973. This standard does not use the term "special items" but rather paragraph 26 states that "A material event or transaction that is unusual in nature or occurs infrequently but not both … should be reported as a separate component of income from continuing operations. The nature and financial effects of each event or transaction should be disclosed on the face of the income statement or, alternatively, in notes to the financial statements." The results in Table 4 suggest that five items account for the majority of both the increase and the level of these separately disclosed income components: asset write-downs, sales of assets, restructurings, and, toward the end of the period, intangibles impairment and debt extinguishment. We examine specific accounting standards that were adopted during the study period that apply to each of these items.

Our search reveals that the earliest accounting standard related to special items adopted during our study period is Staff Accounting Bulletin 67 on restructurings, which was issued by the SEC in 1986 and is discussed in more detail below. However, Figure 1 suggests that the frequency of special items was increasing prior to 1986. To evaluate whether this is the case, we estimate the following time-series regression, with one observation per year from 1975 to 1985:

$$\%SI_t = \gamma_0 + \gamma_1 Trend_t + \gamma_2 E\text{-}ScoreAvg_t + \varepsilon_t, \tag{4}$$

where $\%SI_t$ is the percent of firms reporting a special item in year t, and *Trend* is a variable that goes from 1 (for 1975) to 11 (for 1985). $E\text{-}ScoreAvg_t$ is the average for *E-Score* (defined earlier) across firms for year t. We estimate Equation (4) using *Accounting Trends and Techniques* data.[17] We compute *%SI* as the sum of the percent of gain special items and loss special items reported in Panels A and B of Table 4 (e.g., for 1983 we use 27 + 26 = 53 percent). As the first column of Table 6 demonstrates, the estimated coefficient for *Trend* is 0.012 ($t = 7.67$; $p < 0.01$) and the estimated coefficient for *E-ScoreAvg* is 0.032 ($t = 2.08$; $p = 0.071$). These results indicate a pronounced increase in total special items during the 11-year period prior to the first accounting standard related to special items, as well as some evidence of a role for our economic event proxy, *E-Score*.

Asset Write-Downs

In 1984, the EITF acknowledged that the frequency of asset write-downs had been increasing over time, despite no formal guidance in GAAP for their recognition (see EITF 84-28, in which no consensus on a standard was reached). The first standard to explicitly address asset write-downs was SFAS 121, effective in 1996, with early adoption in 1995 "encouraged." This standard was superseded by SFAS 144 in 2002, but virtually all of the asset impairment rules were left intact. In Table 4, Panel A, asset write-downs increase from 11 percent in 1994 to 19 percent in 1996, but we observe an upward trend both before and after that interval, and larger single year increases in 1998 and 2001. Further, we note that SFAS 121 explicitly cites the widespread reporting of asset impairments without specific authoritative guidance as motivation for the standard (paragraph 5), suggesting this standard was at least partly reactive in nature (see Francis et al. 1996, 117–118).

To evaluate whether SFAS 121 increased asset write-downs, we use data from *Accounting*

[17] Results are very similar and inferences are unchanged using Compustat data for this test. We report results utilizing the *Accounting Trends and Techniques* data for consistency throughout the tests reported in Table 6.

TABLE 6

Regression Analysis
Special Item Incidence and Accounting Standards

	Model 4	Model 5	Model 6	Model 7
Intercept	−0.090***	−0.110***	0.003	−0.352***
	(−4.69)	(−3.44)	(0.08)	(−6.05)
Trend	0.012***	0.006***	0.007***	0.015***
	(7.67)	(3.98)	(5.92)	(5.66)
E-ScoreAvg	0.032*	0.069**	0.101**	0.201***
	(2.08)	(2.29)	(2.29)	(5.19)
D86				−0.019
				(−0.68)
D95		0.118***		−0.037
		(4.78)		(−1.19)
Adj. R^2	0.914	0.910	0.688	0.887
n	11	30	30	26

*, **, *** Represent significance at the 10 percent, 5 percent, and 1 percent levels, respectively, in one-tailed tests (positive coefficients expected), except the intercept.

t-statistics are in parentheses below the coefficient estimates.

This table presents results from estimations of the following models:

Model 4: $\%SI_t = \gamma_0 + \gamma_1 Trend_t + \gamma_2 E\text{-}ScoreAvg_t + \varepsilon_t$.

Model 5: $\%WD_t = \gamma_0 + \gamma_1 Trend_t + \gamma_2 D95_t + \gamma_3 E\text{-}ScoreAvg_t + \varepsilon_t$.

Model 6: $\%SA_t = \gamma_0 + \gamma_1 Trend_t + \gamma_2 E\text{-}ScoreAvg_t + \varepsilon_t$.

Model 7: $\%Rest_t = \gamma_0 + \gamma_1 Trend_t + \gamma_2 D86_t + \gamma_3 D95_t + \gamma_4 E\text{-}ScoreAvg_t + \varepsilon_t$.

The models are OLS time-series regressions with 11–30 annual observations from 1975–2005, as noted under the model number, depending on data availability from *Accounting Trends and Techniques*, summarized in Table 4.

E-ScoreAvg is annual average of the *E-Score* variable defined in Table 5. All main variables are defined in the Appendix.

Trends and Techniques in Table 4 to estimate the following time-series regression, with one observation each year from 1976 to 2005:

$$\%WD_t = \gamma_0 + \gamma_1 Trend_t + \gamma_2 D95_t + \gamma_3 E\text{-}ScoreAvg_t + \varepsilon_t, \tag{5}$$

where $\%WD_t$ is the percent of firms surveyed by *Accounting Trends and Techniques* that reported an asset write-down in year *t*, *Trend* takes values from 1 (for 1976) to 30 (for 2005), and *D95* is an indicator variable that takes the value of 1 if the year is 1995 or later, and 0 otherwise. $E\text{-}ScoreAvg_t$ is as defined earlier.

The second column of Table 6 reports estimates from Equation (5). Consistent with the anecdotal observations in EITF 84-28 and SFAS 121, there is evidence of a strong independent time trend in the incidence of asset write-downs ($t = 3.98$; $p < 0.01$). There is also evidence that our index of economic events helps explain the incidence of write-down activity over time ($t = 2.29$; $p = 0.03$). Finally, the coefficient for *D95* is 0.118 ($t = 4.78$; $p < 0.01$), indicating that the incidence of asset write-downs increased about 12 percentage points in the years following SFAS

121. Thus, it appears that SFAS 121 likely encouraged or legitimized asset write-down activity.

Sale of Assets

Table 4 reports a slow increase over time for this item for both gains and losses. However, because no specific standard applies to asset sales, this contribution does not appear to be due to changes in accounting standards. To evaluate whether the frequency of asset sales is increasing even though it is not affected by a new accounting standard, we estimate the following time-series regression, with one observation per year from 1975 to 2005:

$$\%SA_t = \gamma_0 + \gamma_1 Trend_t + \gamma_2 E\text{-}ScoreAvg_t + \varepsilon_t, \tag{6}$$

where $\%SA_t$ is the percent of firms surveyed by *Accounting Trends and Techniques* that reported a gain or loss from asset sales in year t, and *Trend* is a variable that goes from 1 (for 1975) to 31 (for 2005). $E\text{-}ScoreAvg_t$ is as defined earlier. The third column in Table 6 indicates that the estimated coefficient for *Trend* is 0.007 ($t = 5.92$; $p < 0.01$), consistent with the frequency of gains and losses from asset sales growing significantly during this period when no accounting standards related to asset sales were adopted. In addition, the estimated coefficient for *E-ScoreAvg* is 0.101 ($t = 2.29$; $p = 0.03$), consistent with economic events playing a role in the increase in asset sales.

Restructuring of Operations

The first standard to address this issue is SEC Staff Accounting Bulletin 67, issued in 1986, which states that restructurings should be considered part of operations and disclosed separately if material. Later, EITF 94-3 (adopted in 1995) provides guidance on when a restructuring liability can be recorded. This guidance was revised by SFAS 146 in 2002 (effective in 2003). Restructurings were first reported in *Accounting Trends and Techniques* in 1980 when the frequency was 8 percent. Six years later, the SEC reacted to this new disclosure by issuing SAB 67. Overall, there is a steady increase in the frequency of restructurings throughout the period with spikes in 1986, 1993, 2001, and 2002. One of these is coincident with SAB 67, which may have had the effect of legitimizing this type of disclosure. The spike in 1993 could have resulted from the economic downturn at that time, and may have precipitated the Emerging Issues Task Force to consider the issue in 1994. The additional spikes in 2001 and 2002 are likely the result of 9/11 and the bursting of the technology stock bubble.

To evaluate whether the frequency of special items was affected by these changes in accounting standards, we estimate the following time-series regression, with one observation per year from 1980 to 2005:

$$\%Rest_t = \gamma_0 + \gamma_1 Trend_t + \gamma_2 D86_t + \gamma_3 D95_t + \gamma_4 E\text{-}ScoreAvg_t + \varepsilon_t, \tag{7}$$

where $\%Rest_t$ is the percent of firms surveyed by *Accounting Trends and Techniques* that reported a restructuring loss in year t, *Trend* is a variable that goes from 1 (for 1980) to 26 (for 2005), and *D86* (*D95*) is an indicator variable that takes the value of 1 for all years from 1986 (1995) and later, and 0 (0) otherwise. Thus, *D86* captures the incremental increase in the incidence of restructuring activity from 1986 to 1994 (after SAB 67), and *D95* captures the incremental shift in 1995 and thereafter (after EITF 94-3). $E\text{-}ScoreAvg_t$ is as defined earlier.

In the fourth column of Table 6, the estimated coefficient for *Trend* is 0.015 ($t = 5.66$; $p < 0.01$), indicating a significant increase in the frequency of restructurings. The estimated coefficients on *D86* and *D95* are insignificant, indicating no increase in restructurings after these accounting standards. Finally, the estimated coefficient of 0.201 for *E-ScoreAvg* is significant ($t = 5.19$; $p < 0.01$), consistent with economic activity explaining some of the increasing incidence of restructurings.

Impairment of Intangibles and Debt Extinguishment

Intangible assets were subject to SFAS 121, but *Accounting Trends and Techniques* did not begin reporting this item until 2003, after the adoption of SFAS 142. This standard applies specifically to intangible assets and took effect in 2002. SFAS 145, which took effect in 2003 for calendar-year firms, applies to debt extinguishment transactions. This item first appeared in *Accounting Trends and Techniques* in 2003 at the 10 percent frequency level, and has maintained about that level since. We do not conduct tests for these items because SFAS 142 and SFAS 145 were adopted too recently to explain the nearly 30 years of increasing frequency of special items prior to 2002.

The results summarized above indicate that the incidence of total special items was increasing for some years prior to the adoption of a specific accounting standard addressing special items. We also find that gains and losses from asset sales were increasing throughout the study period even though no new accounting standard affected these transactions. However, there is evidence that a new accounting standard for asset write-downs (SFAS 121) could have contributed to the increasing frequency of these special items. Importantly, our proxy for economic events (*E-ScoreAvg*) is significant in each test.[18] Thus, in general, we observe a stronger and more sustained role for changes in economic activity as an explanation for the increasing frequency of special items and a narrower role for changes in specific accounting standards.

To combine the above analyses into one test, we use Compustat data to compare the relative contributions of specific accounting standards and the economic activity represented by *E-Score* to the increase in the incidence of *total* special items over time. We estimate the following pooled time-series and cross-sectional regression at the firm-year level for 1967–2005:

$$SI_{it} = \gamma_0 + \gamma_1 D86i_{i,t} + \gamma_2 D95_{i,t} + \gamma_3 D96_{i,t} + \gamma_4 D02_{i,t} + \gamma_5 D03_{i,t} + \gamma_6 E\text{-}Score_{i,t} + \gamma_7 Trend_{i,t} + \varepsilon_{i,t} \tag{8}$$

In Equation (8), SI_{it} is an indicator variable equal to 1 if firm i in year t reported a large special item, and 0 otherwise. *D86*, *D95*, and *D96* are defined as in Equations (4)–(7) and *D02* and *D03* are defined analogously (2002 for SFAS 142 and 2003 for SFAS 145 and SFAS 146). Thus, these indicator variables are defined such that successive variables indicate the incremental frequency of special items until the next indicator variable is turned on. For example, the coefficient estimate on *D96* represents the incremental frequency for the period 1996 to 2001 (prior to *D02*), relative to the period prior to 1996. *E-Score* and *Trend* are as defined above.[19]

The results for our estimation of Equation (8) are presented in the first column of Table 7. None of the coefficient estimates for the accounting standard indicator variables are positive. In fact, several are significantly negative.[20] Thus, once the time trend and *E-Score* are controlled, these results provide no evidence that the increase in total special items is due to the implementation of these accounting standards. In contrast, the coefficient estimate on *E-Score* is positive (0.512) and highly significant ($t = 24.87$; $p < 0.01$). The coefficient estimate on *Trend* is also positive (0.127) and highly significant ($t = 7.98$; $p < 0.01$), indicating that, even after controlling for *E-Score*, the incidence of special items is increasing substantially during the study period.

As an alternative analysis, we also estimate the following time-series regression with one observation for each year from 1967–2005:

---

[18] Inferences with respect to accounting standards and the time trend are unchanged if *E-ScoreAvg* is omitted.
[19] We are grateful to an anonymous reviewer for suggesting the analysis in models 8 and 9.
[20] The coefficient estimates for *D02* and *D03* are significantly *negative*, indicating that special items fell significantly from the 2001 level (likely due to 9/11). The coefficient estimate on *D95* is also significantly negative.

TABLE 7

Regression Analysis
Accounting Standards and Economic Events

	Model 8	Model 9
Intercept	−4.583***	−0.073***
	(−20.12)	(−4.06)
D86	−0.173	0.029
	(−1.03)	(1.51)
D95	−0.249**	0.005
	(−2.25)	(0.17)
D96	−0.129	0.046
	(−1.64)	(1.51)
D02	−0.669***	−0.062*
	(−8.25)	(−2.03)
D03	−0.321***	−0.032
	(−4.93)	(−0.97)
E-Score	0.512***	
	(24.87)	
E-ScoreAvg		0.076***
		(3.53)
Trend	0.127***	0.008***
	(7.98)	(7.03)
Pseudo/Adj. R^2	0.147	0.954

*, **, *** Represent significance at the 10 percent, 5 percent, and 1 percent levels, respectively, in one-tailed tests (positive coefficients expected), except the intercept.
t-statistics are in parentheses below the coefficient estimates and are based on standard errors clustered by firm and year.
This table presents results from estimations of the following models:

Model 8: $SI_{i,t} = \gamma_0 + \gamma_1 D86_{i,t} + \gamma_2 D95_{i,t} + \gamma_3 D96_{i,t} + \gamma_4 D02_{i,t} + \gamma_5 D03_{i,t} + \gamma_6 E\text{-}Score_{i,t} + \gamma_7 Trend_{i,t} + \varepsilon_{i,t}$.

Model 9: $\%SI_t = \gamma_0 + \gamma_1 D86_t + \gamma_2 D95_t + \gamma_3 D96_t + \gamma_4 D02_t + \gamma_5 D03_t + \gamma_6 E\text{-}Score_t + \gamma_7 Trend_t + \varepsilon_t$.

Model 8 is a pooled logistic regression at the firm-year level where *SI* is an indicator variable taking the value of 1 if a large special item is reported, 0 otherwise. *E-Score* is the firm-year variable defined in Table 5. The time period is 1967–2005.
Model 9 is OLS time-series regression with 39 annual observations from 1967–2005. In model 9, *E-ScoreAvg* is the annual average of *E-Score*.
Other variables are defined in the Appendix.

$$\%SI_t = \gamma_0 + \gamma_1 D86_t + \gamma_2 D95_t + \gamma_3 D96_t + \gamma_4 D02_t + \gamma_5 D03_t + \gamma_6 E\text{-}ScoreAvg_t + \gamma_7 Trend_t + \varepsilon_t \quad (9)$$

In Equation (9), $\%SI_t$ is the percent of observations with a special item greater than 1 percent of total assets for year t. The year indicator variables in Equation (9) are identical to those in Equation (8), while *E-ScoreAvg* and *Trend* are as defined earlier in Equations (4)–(7).

The results from estimating Equation (9) are presented in the second column of Table 7. Consistent with the results for Equation (8), none of the coefficient estimates on the accounting standard indicator variables are significantly positive. This suggests that these accounting standards did not have a significant effect on the incidence of total special items once the time trend

and *E-ScoreAvg* are controlled. The coefficient estimate on *E-ScoreAvg* is positive (0.076) and significant ($t = 3.53$; $p < 0.01$). This indicates that, after controlling for the time trend, the incidence of special items is 7.6 percent higher for every one-point increase in average *E-Score*.[21]

The results for Equations (8) and (9) provide no support that adoption of the specific accounting standards we investigate contributed significantly to the increase in total special items. In contrast, our composite measure of economic events is significantly related to the incidence of special items over the study period.[22]

VI. COMPETITIVE PRESSURE AND THE CHANGING ECONOMIC ENVIRONMENT

The evidence presented in the previous sections suggests that the increasing incidence of special items in recent decades is attributable, at least in part, to changes in the frequency of economic events associated with recording special items. In this section, we examine whether the increase in this type of economic activity is related to increases in the level of competition in the U.S. (and global) economy over the last 40 years that are documented in previous research.

Prior Research on Increases in Competition

A growing literature provides evidence that the level of competition in the U.S. economy has increased over the last several decades due to changes in trade, technology, and regulation, among other factors. For example, Shepherd (1982) documents a large increase in the level of competition throughout the U.S. economy between 1958 and 1980, which he attributes primarily to antitrust policies, with additional contributions from import competition and deregulation. Pryor (1994) reports that U.S. concentration ratios, a common inverse indicator of competition used in the industrial organization literature, declined between 1958 and 1982 when adjusted for imports, indicating that imports have contributed to increasing competition in U.S. industries.[23]

More recent studies by Comin and Philippon (2005), Gaspar and Massa (2006), and Irvine and Pontiff (2009) document increases in competition later in our study period that they link to increases in firm volatility or performance instability. In addition, Klein and Marquardt (2006) show that the well-documented increasing incidence of accounting losses is attributable mostly to nonaccounting factors such as real firm performance, consistent with an increase in competition over time. In the executive labor market, Huson et al. (2001) provide evidence that forced CEO turnover has increased in recent years, and DeFond and Park (1999) report that the frequency of CEO turnover is positively related to the extent of competition in the industry.

Additional studies provide evidence that firms take actions commonly associated with recording special items in response to competitive pressure. For example, Bernard et al. (2006) report that import-related competition is negatively associated with plant survival and growth, suggesting an association between import competition and asset write-offs. Holmstrom and Kaplan (2001, 121) argue that corporations now regularly transform themselves, through "large waves of merger, takeover and restructuring" that they attribute to "increased dominance of capital markets." Harford (2005) provides evidence that merger waves between 1981 and 2000 were related to deregu-

[21] We also estimate a "changes" version of Equation (9), where the indicator variables for accounting standards include both the implementation year and the following year to capture incremental changes in incidence following the passage of various accounting standards. The change in the percentage of large special items is strongly related to the change in average *E-Score* (Δ*E-ScoreAvg*). The coefficient estimate for Δ*E-ScoreAvg* is 0.072 ($t = 2.86$), the coefficient for time trend is insignificant, and no accounting standard variable is positive and significant.

[22] We view Equations (8) and (9) as complementary. Equation (8) allows *E-Score* to explain cross-sectional variation in the incidence of special items. This may control for "noise" unrelated to new accounting standards, resulting in more powerful tests for the effects of new accounting standards. Alternatively, this may control some of the variation due to new accounting standards, resulting in weaker tests for the effects of accounting standards.

[23] For additional discussion of increasing competition and its possible drivers, see Blinder (2000), Whitman (1999), Reich (2007), and London (2004).

lation and competition among industry peers (see Table 2). Taken together, this research provides direct evidence that competition has increased in the U.S. economy and indirect evidence that this increase in competition is related to company actions that are associated with recognizing special items.

Empirical Analysis

In this subsection, we examine whether increases in competitive pressure can help explain increases in our index of economic events related to special item recognition. Our first measure of competition is the Herfindahl index, a standard measure in industrial organization (e.g., Gaspar and Massa 2006). To compute the Herfindahl Index (*Herf*) we use the Compustat universe to sum across each industry and year each firm's squared market share, where market share is firm revenue divided by industry revenue. We supplement this standard measure of competition with Altman's (1968) z-score (*Bank*), a common measure of bankruptcy risk or financial distress used in accounting and finance research. We view financial distress as a logical byproduct of increased competition where the economic environment is less forgiving of operational errors. We measure z-score at the industry level in the prior fiscal year to better reflect the competitive environment the firm faces rather than the firm's individual financial distress. Measuring z-score at the industry level on a lagged basis also avoids a mechanical relation between components of *E-Score* and the accounting inputs used to calculate z-score.[24]

We examine these two measures individually and in combination. To combine them, we compute a *P-Score* analogous to our *E-Score* by giving a point for each of the two competitive pressure variables above the median value for that variable based on the sample distribution from 1967–1974, and summing the points for each observation. This sum varies from 0 points (below the median for both variables) to two points (above the median for both variables).

If the economic activity captured by *E-Score* is associated with increasing competitive pressure, we expect our measures of competitive pressure to increase over time and to be correlated with *E-Score* in the cross-section. To examine the behavior of our pressure variables over time, we compute the annual average values for *Herf*, *Bank*, and *P-Score* for each of the four sub-periods used previously, as well as the correlation with a time trend. The results are reported in Table 8, Panel A. All three measures are increasing over time (becoming less negative for *Herf* and *Bank*) and all correlations with the time trend are positive and significant ($p < 0.01$).

To provide further evidence on this increasing trend, we divide the sample into three groups on the basis of *P-Score*. The "high" group consists of observations with *P-Score* equal to 2, the "middle" group consists of observations with *P-Score* equal to 1, and the "low" group consists of observations with *P-Score* equal to 0. The last row of Panel A reports that the average incidence of observations in the high *P-Score* group increases from 19 percent in the first sub-period to 45 percent in the last sub-period. Taken together, these results indicate that the annual averages for these pressure variables are increasing during our study period.

To examine whether our measures of competitive pressure are associated with *E-Score* in the cross-section, we compute the average *E-Score* for the three *P-Score* groups for each of the four sub-periods. The results are reported in Table 8, Panel B. The clearest pattern is looking across the columns for each row, where we observe that the average *E-Score* is increasing across *P-Score* groups (i.e., increasing in competitive pressure) in all four sub-periods. These increases are statistically significant for the last three sub-periods ($p < 0.01$). For example, in the period 1997 to 2005 the average *E-Score* is 0.80 for the low *P-Score* group, as compared to 1.20 for the high *P-Score* group.

---

[24] For both *Herf* and *Bank* we assign observations to industries based on SIC codes at the two-digit level.

TABLE 8

Levels of Competitive Pressure Variables over Time

Panel A: Competitive Pressure Variables by Time Period and Correlation with Time Trend

	Mean 67–74	Mean 75–84	Mean 85–96	Mean 97–05	Time Trend	p-value
Level of *Herf*	−0.07	−0.07	−0.06	−0.05	0.70	<0.01
Level of *Bank*	−3.97	−3.56	−3.41	−3.13	0.56	<0.01
Level of *P-Score*	0.95	1.15	1.17	1.40	0.65	<0.01
Incidence of High *P-Score* Observations	0.19	0.27	0.30	0.45	0.63	<0.01

Panel B: Average *E-Score* for *P-Score* Groups

	Low *P-Score*	Middle *P-Score*	High *P-Score*	Diff
1967–1974	0.69	0.73	0.74	0.04
1975–1984	0.78	0.89	0.96	**0.18**
1985–1996	0.86	1.06	1.03	**0.17**
1997–2005	0.80	1.14	1.20	**0.40**
Diff	0.11	**0.41**	**0.46**	

We use two variables to proxy for competitive pressure, the Herfindahl index (*Herf*) and Altman's z-score (*Bank*). See the Appendix for variable definitions. We give each observation in any period one point for each value of *Herf* or *Bank* below the median value for these variables in the period 1967–1974, and then sum points across observations to form an index called *P-Score*. This index can range from 0 to 2 points. We label observations with a value of 0 (roughly bottom 10 percent of the sample) as having a "Low" *P-Score*, observations with a value of 1 as having a "Middle" *P-Score*, and observations with 2 points as having a "High" *P-Score* (roughly top 25 percent of sample).

Panel A plots average levels of *P-Score* and its components over four sub-periods from 1967–2005, along with a correlation of these time-series with a trend variable. For *Herf* and *Bank*, we use cross-sectional medians each year due to the strong skewness of these continuous variables. We also tabulate the incidence of "High" *P-Score* observations over time.

In Panel B, we calculate average values of *E-Score* across the three *P-Score* groups each sub-period. Bolded differences are significant at the 1 percent level (two-sided).

Looking down the columns in Panel B of Table 8, the results also reveal a statistically significant increase in average *E-Score* over time for observations only in the middle and high *P-Score* groups. This indicates that economic events associated with special items have not increased among firms facing low levels of competitive pressure, but have increased for firms facing medium or high levels of economic pressure.[25] In addition, we re-estimate model 8 after adding *P-Score*. In this regression, *P-Score* is positive and significant, suggesting that competitive pressure contributes to the incidence of special items (untabulated). Overall, this evidence indicates that increases in economic events linked to special items are related to increasing competitive pressure in the economy.

[25] In untabulated analysis, we also estimate annual regressions of *E-Score* first on the two economic pressure variables, and then separately on *P-Score*. The averages for the estimated annual slope coefficients are positive and highly significant ($p < 0.01$) for both the components of *P-Score* and for *P-Score* itself.

VII. CONCLUSION

This study examines the reasons for the substantial decline in the contemporaneous association between revenue and expense over time, as documented by Dichev and Tang (2008). To isolate the source of this decline, we first disaggregate total expense into six components. We find that this decline (and the associated increase in earnings volatility and decrease in earnings persistence) is attributable primarily to a steady increase in the frequency of large special items.

We examine two potential reasons for this increase in special items. The first is that changes in specific economic activity have led to more special items. The second is that changes in specific accounting standards require (or encourage) firms to report more special items. We find little evidence that the specific accounting standards we study are a primary reason for the increase in special items. In contrast, we find much stronger evidence that changes in economic activity are a primary reason for the increase in special items. In particular, we find that an index of five indicators of specific economic events associated with recording special items has increased steadily over time and that high values for the index have more than twice the incidence of special items as low index values.

Finally, we examine whether increasing competition in the U.S. economy is related to increases in our index of economic events. We identify two indicators of competitive pressure at the industry level, finding that, both individually and when combined, they are increasing over time and are correlated with our economic activity index in the cross-section.

Our findings demonstrate the importance of disaggregating special items, and are important for standard-setters, given the frequency of these items. This is especially true in light of the recent discussion paper by the FASB and the IASB, *Preliminary Views on Financial Statement Presentation* (FASB 2008), which directly addresses this issue. The FASB's preliminary view is that special items should be disclosed in a "memo" column of the proposed reconciliation between cash flows and comprehensive income (paras. 4.48 to 4.52), while the IASB does not support this additional disclosure (para. 4.53).

While the evidence presented in this study suggests that changes in economic activity are a primary reason for the increase in special items and specific accounting standards contribute little to the increase in special items, our results do not prove the null hypothesis that accounting standards play no role in the increase in special items. The inability of our economic variables to fully explain the time trend in special items leaves open the possibility that accounting standards could explain some of this unexplained variation. Thus, it is possible that future research will develop stronger tests that are able to detect a role for accounting standards.

We also leave two additional important issues for future research. The first is the possibility that changes in financial reporting practice that are not related to either new accounting standards or to changes in economic activity have contributed to the increase in special items over time. For example, perhaps managers facing increasing capital market pressure have changed their accounting practices and are now more aggressively recognizing nonrecurring gains and losses, encouraging investors to ignore these items for performance evaluation. This issue is beyond the scope of this study. Also beyond the scope of this study is whether performance measures that so frequently include special items are "good accounting." While our evidence indicates that the increase in special items over time is linked to increases in economic events that affect firm welfare, we leave a more complete evaluation of the appropriateness of this accounting to future research.

APPENDIX
DETAILED VARIABLE DEFINITIONS

Earnings Variables and Components

Revenue (*Rev*) = sales (data12)/average total assets (data6);
Expense (*Exp*) = sales less income before extraordinary items (data18)/ average total assets (data6);
Cost of Goods Sold (*COGS*) = data 41/average total assets (data6);
Selling, General, and Admin (*SGA*) = data189/average total assets (data6);
Depreciation (*DEPR*) = data 14/average total assets (data6), equals 0 if missing;
Tax Expense (*TAX*) = data16/average total assets (data6);
Special Items (*SI*) = $(-1) * $ data17/average total assets (data6) for decomposition analysis, equals 0 if missing. An indicator variable equal to 1 if the absolute value of special items is greater than or equal to 1 percent of average total assets (data6), 0 otherwise. Cross-sectional average measures annual incidence;
Operating Income after Depreciation = data178/average total assets (data6);
Other Expenses (*OTH*) = total expense less *COGS*, *SGA*, *DEPR*, *TAX*, and *SI*; and
$\%SI_t$ = percent of sample firm-year observations with a special item greater than 1 percent of beginning total assets that year.

Economic Event Variables

Negative Employee Growth (*Emp*) = indicator variable that takes the value of 1 if the firm had negative employee (data29) growth for the year relative to the preceding year;
Merger and Acquisition Activity (*MA*) = indicator variable that takes the value of 1 if the firm engaged in a merger or acquisition during the fiscal year (from AFTNT1);
Discontinued Operations (*Disc*) = indicator variable that takes the value of 1 if the firm had discontinued operations for the year (data66);
Operating Loss (*Loss*) = indicator variable that takes the value of 1 if the firm had negative operating income after depreciation (data178) for the year;
Negative Revenue Growth (*Grow*) = indicator variable that takes the value of 1 if the firm had negative revenue (data12) growth rate, 0 otherwise;
E-Score = sum of the five indicator variables *Emp*, *MA*, *Disc*, *Loss*, and *Grow*; and
E-ScoreAvg = cross-sectional average of *E-Score* for a given year.

Accounting Standard Variables

Dxx = 1 in 19xx or 20xx and beyond, and 0 prior to 19xx or 20xx.

Competitive Pressure Variables

Herfindahl Index (*Herf*) = sum for each industry-year of member firm's squared market share. Market share is firm revenue divided by industry revenue using all available Compustat data;

Altman's Bankruptcy Prediction (*Bank*) = (−1) × the industry median z-score from Altman (1968) from the prior year, so that *Bank* is increasing in competition, varies from the lowest probability of bankruptcy (the lowest scores) to the highest probability of bankruptcy (the highest scores); and

P-Score = sum of a point for each of *Herf* and *Bank* that are above the sample median from 1967–1974.

REFERENCES

Altman, E. I. 1968. Financial ratios, discriminant analysis, and the prediction of corporate bankruptcy. *The Journal of Finance* 23 (4): 589–609.

American Institute of Certified Public Accountants. (AICPA). 1973. *Reporting the Results of Operations—Reporting the Effects of Disposal of a Segment of a Business, and Extraordinary, Unusual, and Infrequently Occurring Events and Transactions*. Accounting Principles Board Opinion No. 30. New York, NY: AICPA.

Bernard, A., J. Jensen, and P. Schott. 2006. Survival of the best fit: Exposure to low-wage countries and the (uneven) growth of U.S. manufacturing plants. *Journal of International Economics* 68 (1): 219–237.

Blinder, A. 2000. How the economy came to resemble the model. *Business Economics (Cleveland, Ohio)* (January): 16–25.

Burgstahler, D., J. Jiambalvo, and T. Shevlin. 2002. Do stock prices fully reflect the implications of special items for future earnings? *Journal of Accounting Research* 40 (3): 585–612.

Collins, D., E. Maydew, and I. Weiss. 1997. Changes in the value-relevance of earnings and book values over the past forty years. *Journal of Accounting and Economics* 24 (1): 39–67.

Comin, D., and T. Philippon. 2005. *The Rise in Firm-Level Volatility: Causes and Consequences*. Cambridge, MA: NBER Macroannuals.

Cready, W., T. J. Lopez, and C. A. Sisneros. 2010. The persistence and market valuation of recurring nonrecurring items. *The Accounting Review* 85 (5): 1577–1615.

Dechow, P. M. 1994. Accounting measures of earnings and cash flows as measures of firm performance: The role of accounting accruals. *Journal of Accounting and Economics* 18 (1): 3–42.

———, and W. Ge. 2006. The persistence of earnings and cash flows and the role of special items: Implications for the accrual anomaly. *Review of Accounting Studies* 11 (2–3): 253–296.

———, and C. M. Schrand. 2004. *Earnings Quality*. Charlottesville, VA: The Research Foundation of CFA Institute.

DeFond, M., and C. Park. 1999. The effect of competition on CEO turnover. *Journal of Accounting and Economics* 27 (1): 35–56.

Dichev, I., and V. Tang. 2008. Matching and the changing properties of accounting earnings over the last 40 years. *The Accounting Review* 83 (6): 1425–1460.

Elliott, J., and D. Hanna. 1996. Repeated accounting write-offs and the information content of earnings. *Journal of Accounting Research* 34 (Supplement): 135–155.

Fairfield, P., K. Kitching, and V. Tang. 2009. Are special items informative about future profit margins? *Review of Accounting Studies* 14 (2–3): 204–236.

Fan, Y., A. Barua, W. Cready, and W. Thomas. 2010. Managing earnings using classification shifting: Evidence from quarterly special items. *The Accounting Review* 85 (4): 1303–1323.

Financial Accounting Standards Board (FASB). 2008. *Preliminary Views on Financial Statement Presentation*. Norwalk, CT: FASB.

Francis, J., D. Hanna, and L. Vincent. 1996. Causes and effects of discretionary asset write-offs. *Journal of Accounting Research* 34 (Supplement): 117–134.

———, R. LaFond, P. M. Olsson, and K. Schipper. 2004. Costs of equity and earnings attributes. *The Accounting Review* 79 (4): 967–1010.

Frankel, R. 2009. Discussion of: Are special items informative about future profit margins? *Review of Accounting Studies* 14 (2–3): 237–245.

Gaspar, J., and M. Massa. 2006. Idiosyncratic volatility and product market competition. *The Journal of Business* 79 (6): 3125–3152.

Harford, J. 2005. What drives merger waves? *Review of Financial Studies* 77 (September): 529–560.

Holmstrom, B., and S. N. Kaplan. 2001. Corporate governance activity in the United States: Making sense of the 1980s and 1990s. *The Journal of Economic Perspectives* 15 (2): 121–144.

Huson, M., R. Parrino, and L. Starks. 2001. Internal monitoring mechanisms and CEO turnover: A long-term perspective. *The Journal of Finance* 56 (6): 2265–2297.

Irvine, P., and J. Pontiff. 2009. Idiosyncratic return volatility, cash flows, and product market competition. *Review of Financial Studies* 22 (3): 1149–1177.

Kee, H. L. 2009. The coefficient of the whole is an average of the coefficients of its parts: A special case of restricted least squares. *Economics Letters* 104 (3): 136–139.

Klein, A., and C. Marquardt. 2006. Fundamentals of accounting losses. *The Accounting Review* 81 (1): 179–206.

London, P. 2004. *Competition Solution: The Bipartisan Secret behind American Prosperity*. Washington, D.C.: AEI Press.

McVay, S. 2006. Earnings management using classification shifting: An examination of core earnings and special items. *The Accounting Review* 81 (3): 501–532.

Pryor, F. 1994. The evolution of competition in U.S. manufacturing. *Review of Industrial Organization* 9 (6): 695–715.

Reich, R. 2007. *Supercapitalism: The Transformation of Business, Democracy, and Everyday Life*. New York, NY: Alfred A. Knopf.

Shepherd, W. G. 1982. Causes of increased competition in the U.S. economy, 1939–80. *The Review of Economics and Statistics* 64 (4): 613–626.

Standard & Poor's. 2003. *Compustat User's Guide*. Centennial, CO: The McGraw-Hill Companies, Inc.

Whitman, M. N. 1999. *New World, New Rules: The Changing Role of the American Corporation*. Cambridge, MA: Harvard Business Press.

THE ACCOUNTING REVIEW
Vol. 86, No. 3
2011
pp. 975–1005

American Accounting Association
DOI: 10.2308/accr.00000041

Internal Controls and Conditional Conservatism

Beng Wee Goh
Singapore Management University

Dan Li
Tsinghua University

ABSTRACT: This study examines the relation between internal controls and conditional conservatism ("conservatism"), also referred to as timely loss recognition. Using a sample of firms that disclose material weaknesses (MWs) in internal controls under the Sarbanes-Oxley Act (SOX), we find a positive relation between internal control quality and conservatism. Specifically, firms with MWs exhibit lower conservatism than firms without such weaknesses. Further, firms that disclose MWs and subsequently remediate these weaknesses exhibit greater conservatism than firms that continue to have MWs. Overall, these results are consistent with strong internal controls acting as a mechanism that facilitates conservatism. Our study contributes to the literature on the reporting effects of strong versus weak internal controls.

Keywords: *internal controls; conservatism; material weaknesses; Sarbanes-Oxley Act.*

Data Availability: *Data are available from sources identified in the text.*

I. INTRODUCTION

In 2002, the U.S. Congress passed the Sarbanes-Oxley Act (SOX) to improve the quality of financial reporting and to restore investor confidence in the reliability of financial statements. An important aspect of SOX is its internal control reporting requirements, which allow investors to be informed about the quality of a firm's internal controls. Specifically, Section 302 of the Act requires management to evaluate and report the effectiveness of disclosure and control procedures (SEC 2002), while Section 404 requires the auditor to form an opinion on the effective-

We thank Steve Kachelmeier (senior editor) and two anonymous referees for suggestions and guidance that have significantly improved this paper. We also thank Bryan Church, Arnold Schneider, Eugene Comiskey, Steve Matsunaga, Jayanthi Krishnan, Jeffrey Ng, Michael Willenborg, Patricia Dechow, Yuan Zhang, Zhaoyang Gu, Ole-Christian Hope, Stephen Penman, Chee Yeow Lim, Chih-Ying Chen, and participants at the 2008 AAA Annual Meeting for their comments. We also appreciate the financial support from Tsinghua University and funding through a research grant from the Office of Research, Singapore Management University.

Editor's note: Accepted by Steven Kachelmeier.

Submitted: January 2009
Accepted: October 2010
Published Online: April 2011

ness of internal controls (SEC 2003).[1] Using these internal control disclosures, studies have investigated whether the quality of internal controls affects accruals quality (Doyle et al. 2007a; Ashbaugh-Skaife et al. 2008), a firm's cost of capital (Ogneva et al. 2007; Beneish et al. 2008; Ashbaugh-Skaife et al. 2009), and the accuracy of management guidance (Feng et al. 2009). Given the importance of the internal control provisions as a means to improve the governance of firms, our study extends the literature on the reporting effects of strong versus weak internal controls by examining how the quality of internal controls is related to conservatism in financial reporting.

We focus on conservatism because it has been argued to provide several governance benefits, such as reducing agency conflicts and improving managerial investment decisions (Holthausen and Watts 2001; Watts 2003; Ball and Shivakumar 2005), enhancing the efficiency of debt contracts (Ahmed et al. 2002; Zhang 2008), and reducing litigation costs (Watts 2003). Following Basu (1997), we define conservatism as the higher degree of verification to recognize good news as gains than to recognize bad news as losses. That is, earnings reflect bad news more quickly than good news leading to timelier loss recognition. The literature generally refers to this as conditional conservatism (Beaver and Ryan 2005; Ball and Shivakumar 2005), which we simply refer to as "conservatism" in this study.[2]

Within the agency framework of positive accounting theory that recognizes the contracting benefits of conservatism, it is plausible to expect that strong internal controls facilitate conservative financial reporting. First, firms with a strong internal control environment (e.g., strong tone-at-the-top and/or good internal control culture) are more likely to understand the role of conservatism in contracting and in reducing agency conflicts. Consequently, these firms are more likely to favor its implementation. Further, to the extent that firms are committed to produce conservative reports, strong internal controls can facilitate this process by providing reliable accounting information. Weak internal controls can cause estimation errors in accounting numbers (Doyle et al. 2007a; Ashbaugh-Skaife et al. 2008), making the contractible variables less reliable for contracting and monitoring purposes. This can reduce the effectiveness of conservatism as a governance mechanism and, hence, the incentives to implement conservatism. Finally, weak internal controls could impede the timely recognition of losses, thereby leading to lower conservatism. We develop these arguments on the positive relation between internal control quality and conservatism more fully in Section II.

Moving away from the agency framework of positive accounting theory that argues the contracting benefits of conservatism, it becomes less clear why one would expect a positive relation between internal control quality and conservatism. First, it has been argued that conservatism can increase earnings management (Levitt 1998; Penman and Zhang 2002), which can exacerbate agency conflicts. Under this view, if strong internal controls serve as an effective monitoring mechanism that can mitigate agency problems (Jensen 1993), then we should instead expect strong internal controls to provide a disincentive for firms to produce conservative reports. Second, it is controversial whether conservatism is a desirable attribute of financial reporting. Barth (2008) argues that conservatism is not a qualitative characteristic under the conceptual framework of IASB, which specifies that accounting information should be unbiased. Conservatism, by promoting the deliberate understatement of book value and/or earnings, would render

---

[1] In Endnote 59 of the SEC's (2002) release on the *Final Rule: Certification of Disclosure in Companies' Quarterly and Annual Reports*, internal control is defined as "a process, effected by an entity's board of directors, management and other personnel, designed to provide reasonable assurance regarding the reliability of financial reporting."

[2] Another form of reporting conservatism, referred to as "unconditional conservatism," manifests itself in a systematic undervaluation of the firm's net assets. Ball and Shivakumar (2005) argue that unconditional conservatism cannot yield contracting efficiency because it can be observed and, hence, can be "undone" by users of the financial statements. Accordingly, we do not consider unconditional conservatism in our context.

financial statements not neutral. Under this investor value relevance framework that emphasizes that accounting numbers should be neutral inputs to investment decisions, the argument that strong internal controls would promote conservative accounting seems tenuous.

Given the above arguments and corresponding tension, we empirically test the relation between internal control quality and conservatism using firms' internal control disclosures under SOX. We obtain a sample of firms that disclosed at least one material weakness (MW) from January 2003 to November 2005 and deem these firms (hereafter referred to as "MW firms") to have low internal control quality due to the presence of MWs.[3] We apply three measures of conservatism that are commonly used in the literature to capture the asymmetric timeliness in the recognition of economic losses: (1) the persistence of earnings changes measure in Basu (1997), (2) the accrual-based loss recognition measure in Ball and Shivakumar (2005), and (3) the timeliness of earnings to news measure in Basu (1997). To afford stronger inferences to be made about the effects of internal controls on conservatism, we further conduct intertemporal tests of changes in the status of internal controls. Specifically, we examine whether MW firms that subsequently remediate their weaknesses exhibit different levels of conservatism from MW firms that continue to have these weaknesses.

Results using all three measures of conservatism are consistent with a positive relation between internal control quality and conservatism. Specifically, we find that MW firms exhibit lower conservatism than control firms without such weaknesses. Further, we find that MW firms that subsequently remediate their weaknesses (i.e., show an improvement in internal control quality) exhibit greater conservatism than MW firms that continue to have these weaknesses. These results are robust to a battery of checks, which include controlling for the self-selection bias in the MW firms sample, controlling for the self-selection bias in the improvement of internal control quality, and using alternative conservatism measures suggested by Khan and Watts (2009) and Givoly and Hayn (2000). Taken together, our results are consistent with strong internal controls serving as a mechanism that facilitates conservatism.

Doyle et al. (2007a) and Ashbaugh-Skaife et al. (2008) find a positive relation between internal control quality and accruals quality, which is measured by the extent to which accruals are realized as cash flows or the size of abnormal accruals. Our study differs from their studies by examining conservatism, which is a form of accounting bias toward reporting low book values of stockholder equity, *conditional on firms experiencing contemporaneous economic losses* (Ball and Shivakumar 2005).[4] In both Doyle et al. (2007a) and Ashbaugh-Skaife et al. (2008), the authors conjecture that their findings could be due to weak internal controls (1) causing unintentional (poor estimation ability) errors, or (2) failing to limit intentional income-increasing earnings management. However, Ashbaugh-Skaife et al. (2008) find that firms that report internal control weaknesses have significantly larger positive and larger negative abnormal accruals relative to

[3] According to Auditing Standards No. 2 (PCAOB 2004), which was in effect during the time period of our study, an MW is "a significant deficiency, or combination of significant deficiencies, that results in more than a remote likelihood that a material misstatement of the financial statements will not be prevented or detected." A significant deficiency is "a control deficiency, or combination of control deficiencies, that adversely affects the company's ability to initiate, authorize, record, process, or report external financial data reliably in accordance with generally accepted accounting principles such that there is more than a remote likelihood that a misstatement of the company's annual or interim financial statements that is more than inconsequential will not be prevented or detected."

[4] Ball and Shivakumar (2006) argue that conventional accruals models assume that the relation between accruals and cash flows are linear and do not take into account of the asymmetry in the gain and loss recognition role of accruals. Hence, these models misspecify the accounting accrual process and misestimate discretionary accruals, resulting in incorrect inferences about earnings quality. By examining the timely loss recognition property of earnings, our study takes into consideration the gain and loss recognition role of accruals and the nonlinearity implied by the asymmetry therein. This addresses the concerns in the use of the linear accrual models.

control firms, which suggests that the first factor is more plausible. Nonetheless, these two factors need not necessarily result in lower conservatism per se.

First, the unintentional estimation errors caused by weak internal controls should affect the recognition of bad news and good news in an unbiased and symmetrical manner. Accordingly, these estimation errors are not likely to cause conservatism per se, which results from the asymmetric response of earnings to bad versus good economic news. Second, even if weak internal controls fail to limit earnings management, the presence of income-increasing earnings management is not necessarily consistent with the presence of lower conservatism. Givoly et al. (2010, 221) argue that earnings management is situational or episodic, occurring when unmanaged earnings fail to meet an important reporting objective (e.g., meeting analysts' forecasts). Further, earnings management tends to be small-scale in nature due to the short supply of the needed positive accruals. Hence, if the company's accounting is generally conservative, earnings management will only temporarily interrupt this observed reporting pattern but will not obscure the presence of the more prevalent phenomenon of conservatism. The implication is that earnings management and conservatism can co-exist within a firm.[5] Hence, our study complements Doyle et al. (2007a) and Ashbaugh-Skaife et al. (2008) by providing insights into how internal control quality is associated with the phenomenon of conservatism.

Overall, our study extends the literature on the reporting effects of strong versus weak internal controls (e.g., Doyle et al. 2007a; Ogneva et al. 2007; Ashbaugh-Skaife et al. 2008; Beneish et al. 2008). We show that strong internal controls can create an incentive for and facilitate conservative reporting. Our study also contributes to the literature that shows that conservatism is shaped, in part, by a firm's contracting and governance environment (e.g., Ball et al. 2000; Basu et al. 2001; Ball et al. 2003; Ball and Shivakumar 2005; Bushman and Piotroski 2006; Ahmed and Duellman 2007; LaFond and Roychowdhury 2008).

Section II develops the hypotheses, Section III explains the research design, Section IV presents the empirical results, Section V describes additional analyses, and Section VI concludes.

II. HYPOTHESIS DEVELOPMENT

In an attempt to restore investor confidence in firms' financial reporting, SOX requires firms to assess and disclose, and auditors to certify, the effectiveness of internal controls over financial reporting (SEC 2002, 2003). Presumably, regulators hope that these requirements can improve the quality of internal controls and enhance the reliability of financial reporting. Indeed, studies have shown that effective internal controls can enhance financial reporting quality, proxied by accruals quality and the size of abnormal accruals (Doyle et al. 2007a; Ashbaugh-Skaife et al. 2008). Further, Beneish et al. (2008) and Ashbaugh-Skaife et al. (2009) find that effective internal controls result in reduced information risk, which can lower a firm's cost of equity. However, Ogneva et al. (2007) fail to find a significant relation between internal control weaknesses and cost of equity after controlling for primitive firm characteristics and analyst forecast bias. Finally, Feng et al. (2009) find a positive relation between internal control quality and the accuracy of management guidance, consistent with ineffective internal controls causing errors in internal management reports. Our study extends this stream of research by examining how the quality of internal controls affects conservatism in financial reporting.

Conservatism has been argued to play an important governance role in mitigating agency conflicts and enhancing contracting efficiency within the firm (Holthausen and Watts 2001; Watts

---

[5] Consistent with this line of reasoning, Givoly et al. (2010) find that public equity firms report more conservatively than their private equity counterparts, but at the same time, public equity firms have a greater propensity to manage income. The authors also conduct two tests to show that earnings management and conservative reporting may independently exist in their data.

2003; Ball and Shivakumar 2005). The reasoning is that conservatism reduces managers' ability and incentives to overstate earnings and net assets by requiring higher verification standards for gain recognition and reduces managers' ability to withhold information on expected losses. Thus, it prevents overcompensation of managers that is costly to recover *ex post* because of managers' limited liability and tenure.[6] Similarly, in a debt-contracting setting, conservatism reduces managers' ability to loosen or avoid dividend restrictions and transfer wealth from bondholders to shareholders, thereby mitigating deadweight losses and increasing firm value (Ahmed et al. 2002; Zhang 2008). According to Ball and Shivakumar (2005), conservative accounting effectively limits the control rights of managers and transfers those rights back to the providers of finance earlier. Finally, Watts (2003) asserts that conservatism can also reduce a firm's expected litigation costs.

Existing studies suggest that the incentives for or against conservative reporting are shaped, in part, by a firm's contracting and governance environment (e.g., Ball et al. 2000; Basu et al. 2001; Ball 2001; Ball et al. 2003; Ball and Shivakumar 2005; Bushman and Piotroski 2006; Ahmed and Duellman 2007; LaFond and Roychowdhury 2008; LaFond and Watts 2008; Nikolaev 2010). For instance, Ahmed et al. (2002) argue that conservatism evolves as an efficient contracting mechanism to mitigate dividend policy conflicts between shareholders and bondholders. They find that the more severe such conflicts are (i.e., the higher the leverage), the more conservative the firm's accounting choices become. LaFond and Roychowdhury (2008) hypothesize and find that when managerial ownership is low, the severity of the agency problem increases, driving a demand for greater conservatism. Ball and Shivakumar (2005) find that private firms are less conservative than public firms and attribute this finding partly to the differences in governance structures and monitoring mechanisms between these two types of firms. Given that the internal control system plays an important governance role within the firm to monitor managers' behavior and to minimize agency costs (Fama and Jensen 1983; Jensen 1993), we conjecture a potential link between the quality of internal controls and conservative reporting.

Within the agency framework of positive accounting theory that recognizes the contracting benefits of conservatism, strong internal controls could therefore create an incentive for conservative accounting reports. According to the *Internal Control—Integrated Framework* (Committee of Sponsoring Organizations of the Treadway Commission 1992), the internal control environment is an important component of internal controls. We argue that firms with a strong tone-at-the-top (e.g., an effective board of directors that protects the interests of shareholders) and good internal control culture (e.g., a corporate culture that has a wealth-maximization orientation) are better able to understand the benefits of conservatism in reducing agency conflicts and deadweight loss, in contracting with shareholders and/or debtholders, and in reducing the litigation risks for directors, auditors, and managers. Consequently, these firms are more likely to favor the implementation of conservatism and promote an environment that emphasizes the adherence to conservative reporting.

To the extent that firms understand the benefits of conservatism and are committed to produce conservative reports, strong internal controls can facilitate the use of conservatism in the contracting process by providing reliable accounting information. This is because the financial reporting system, by emphasizing verifiable outcomes, supplies a rich set of variables that can be used for contracting purposes (Watts and Zimmerman 1986; Bushman and Smith 2001; Armstrong et al. 2010). For instance, financial covenants used in debt contracts to limit managerial discretion with

[6] Conservative accounting reduces the tendency of managers with short-term horizons to invest in negative Net Present Value (NPV) projects, making managers aware that they will not be able to defer the recognition of losses to the future (Ball and Shivakumar 2005) and imposing greater costs to biasing financial reports upward (Guay and Verrecchia 2006).

respect to dividend distributions (Ahmed et al. 2002) are often defined in terms of accounting numbers. Similarly, managers' compensation can be tied to changes in book value (or earnings).

As mentioned earlier, weak internal controls can cause unintentional errors in accrual estimations to occur and impact the reported book value or earnings. For instance, the failure to have appropriate reconciliations and reviews in place allow for procedural errors in accrual estimations (Doyle et al. 2007a), and the lack of adequate policies, training, or diligence by company employees can affect the noise and magnitude of abnormal accruals (Ashbaugh-Skaife et al. 2008). Although these estimation errors cannot cause lower conservatism per se, they can make the contractible variables noisy and less reliable, providing a poor signal to the board of directors that conservatism has been used effectively in contracting and reducing the usefulness of conservatism in monitoring managers.[7] In sum, by enhancing the ability to use accounting numbers effectively in contracts, strong internal controls can create an incentive for firms to use conservatism as a governance mechanism.

Finally, weak internal controls could also impede the timely recognition of losses, thereby leading to lower conservatism. For instance, the lack of proper policies and procedures to regularly value a firm's inventory or fixed assets prevents asset impairment from being discovered early and, hence, delays the recognition of losses. Unqualified accounting staff may lack the expertise to estimate the future cash flows of assets such as goodwill and fixed assets, and identify any reduction in future cash flows that would trigger early recognition of impairment losses. Weak internal controls in the monitoring of the firm's investment decisions can also provide the incentive for managers to avoid timely loss recognition in investment projects. One example is the lack of a proper process to provide the board of directors with timely information on the firm's investments for oversight and review, which could make it easier for managers to invest in or avoid divesting negative NPV projects. Further, control weaknesses over the computation of the present value of future profits of acquired businesses can result in potential loss-making projects not being discovered and terminated early. The Appendix provides some examples of MWs disclosed by firms that could potentially impede the timely recognition of losses.

The above arguments suggest a positive effect of conservative accounting reports in terms of contracting efficiency. However, moving away from the agency framework of conservatism as a beneficial governance mechanism, it becomes less clear why one would expect a positive link between internal controls and conservatism. First of all, it is questionable whether conservatism is beneficial to firms and their shareholders. For instance, Penman and Zhang (2002) contend that conservative accounting can create unrecorded reserves that provide managers with the flexibility to report more income in the future. The increased ability to manage earnings can adversely affect the predictive ability of current earnings for future earnings. In addition, three of the five financial reporting problems highlighted by former SEC Chair Arthur Levitt deal with the understatement of assets—big bath charges, creative acquisition accounting, and cookie jar reserves (Levitt 1998). If so, then conservative accounting reporters may actually have greater ability to smooth or manipulate earnings, which could exacerbate, rather than mitigate, agency problems. If strong internal controls serve as an effective monitoring mechanism to reduce agency conflicts (Jensen 1993),

---

[7] One may question why a firm would not ensure that internal controls are effective or remediate internal control weaknesses promptly if the firm views strong internal controls as important in ensuring conservative reports. This could be due to several reasons. First, the firm may fail to constantly invest in internal controls to keep up with the growth of the firm. Second, managers may not be willing to invest in the time and resources to ensure effective internal controls because such actions divert attention away from the core business. Third, firms may face strained financial resources. Fourth, firms may have complex operations, such as multiple operating segments, making it difficult to implement effective internal controls.

then the claim that strong internal controls promote conservatism would become untenable. Instead, strong internal controls could provide no incentive or even a disincentive for firms to produce conservative accounting reports.

Further, regulators and standard-setters do not necessarily view conservatism as a desirable attribute of financial reporting. For instance, Barth (2008) points out that conservatism is not a qualitative characteristic of accounting information under the conceptual framework of the International Accounting Standards Board (IASB). The framework specifies that accounting information should be unbiased, and freedom from bias is an essential characteristic of reliability and of neutrality (IASB 2001, para. 31, 36). Conservatism, by allowing the deliberate understatement of assets or income and/or the deliberate overstatement of liabilities or expenses, renders financial statements not neutral and, therefore, jeopardizes the quality of reliability and unbiasedness desired by the conceptual framework.

The Financial Accounting Standards Board (FASB) asserts that the best interest of users is served by neutral reporting accompanied by appropriate disclosure of "the nature and extent of the uncertainty surrounding events and transactions reported to stockholders and others" (see SFAC No. 2, FASB 1980, para. 93; FASB 2000, para. 96, 97).[8] Consistent with this perspective, the FASB is shifting its focus toward fair value accounting, lessening the asymmetric treatment of bad and good news in financial statements. If the objective of financial reporting is to provide information useful for making economic decisions (IASB 2001, para. 12), then having strong internal controls that ensure that the financial reports are unbiased would satisfy investors' demand for neutral information. Hence, within this investor value-relevance framework, the argument that strong internal controls would promote conservative financial reports (which are essentially biased) seems tenuous.

Given these potentially conflicting perspectives, we empirically test the relation between internal controls and conservatism and test the null hypothesis that internal control quality is unrelated to conservatism. In order to further ascertain the nature of any relation between internal controls and conservatism, we perform intertemporal tests, as in Ashbaugh-Skaife et al. (2008), to determine whether firms with weak internal controls that subsequently show an improvement in internal controls exhibit different levels of conservatism from firms that fail to do so. This would help strengthen our finding on the relation between internal controls and conservatism. The null hypotheses we test are as follows:

H1: There is no relation between internal control quality and conservatism.

H2: Firms with weak internal controls that subsequently show an improvement in internal controls (i.e., remediate their internal control weaknesses) exhibit no difference in conservatism from firms that fail to do so.

III. RESEARCH DESIGN

Measures of Accounting Conservatism

Persistence of Earnings Changes (Basu 1997)

Our first measure to capture the differential timeliness of loss versus gain recognition is the persistence of earnings changes in Basu (1997). Basu (1997) shows that, relative to good news

[8] For example, in SFAC No. 2 the FASB states, "Conservatism in financial reporting should no longer connote deliberate, consistent understatement of net assets and profits" (FASB 1980, para. 93). The FASB also states in the same paragraph that conservatism "became deeply ingrained and is still in evidence despite efforts over the past 40 years to change it." In the joint exposure draft by the FASB and IASB, *Conceptual Framework for Financial Reporting* (FASB 2008, para. BC2.21), it states that "the boards concluded that describing *prudence* or *conservatism* as a qualitative characteristic or a desirable response to uncertainty would conflict with the quality of neutrality because … an admonition to be prudent is likely to lead to a bias in the reported financial position and financial performance … Accordingly, the proposed framework does not include *prudence* or *conservatism* as desirable qualities of financial reporting information" (emphasis in the original).

periods, conservatism results in lower persistence of earnings in bad news periods. The deferred recognition of relatively good news results in positive changes in income being less likely to reverse than negative earnings changes. This is because, from a time-series perspective, the bad news reflected in current earnings will appear as a transitory shock in the earnings process, whereas the effects of a current positive shock will be spread over the earnings of several future periods as anticipated gains are realized. The following model from Basu· (1997) is used to estimate this relation:

$$\Delta NI_t = \alpha_0 + \alpha_1 D\Delta NI_{t-1} + \alpha_2 \Delta NI_{t-1} + \alpha_3 D\Delta NI_{t-1} * \Delta NI_{t-1} + \varepsilon \tag{1}$$

where firm i subscripts are omitted. $\Delta NI_t(\Delta NI_{t-1})$ is the change in net income before extraordinary items for firm i in fiscal year t ($t-1$) deflated by beginning-of-year total assets, and $D\Delta NI_{t-1}$ is an indicator variable that equals 1 if ΔNI_{t-1} is less than 0, and 0 otherwise.

The negative coefficient on $D\Delta NI_{t-1} * \Delta NI_{t-1}$ is consistent with timely loss recognition. We further develop model (1) into model (2) to test H1, where MW is an indicator variable that equals 1 if the firm has at least one MW in internal controls, and 0 otherwise. If MW firms report less (more) conservatively than firms without such weaknesses, then the coefficient on $D\Delta NI_{t-1} * \Delta NI_{t-1} * MW$ will be positive (negative); that is, MW firms have a reduced (increased) tendency to reverse negative earnings changes in the following period.

$$\begin{aligned}\Delta NI_t = {} & \alpha_0 + \alpha_1 D\Delta NI_{t-1} + \alpha_2 \Delta NI_{t-1} + \alpha_3 D\Delta NI_{t-1} * \Delta NI_{t-1} + \alpha_4 MW + \alpha_5 D\Delta NI_{t-1} * MW \\ & + \alpha_6 \Delta NI_{t-1} * MW + \alpha_7 D\Delta NI_{t-1} * \Delta NI_{t-1} * MW + \alpha_8 MB_t + \alpha_9 D\Delta NI_{t-1} * MB_t \\ & + \alpha_{10} \Delta NI_{t-1} * MB_t + \alpha_{11} D\Delta NI_{t-1} * \Delta NI_{t-1} * MB_t + \alpha_{12} LEV_t + \alpha_{13} D\Delta NI_{t-1} * LEV_t \\ & + \alpha_{14} \Delta NI_{t-1} * LEV_t + \alpha_{15} D\Delta NI_{t-1} * \Delta NI_{t-1} * LEV_t + \alpha_{16} SIZE_t + \alpha_{17} D\Delta NI_{t-1} * SIZE_t \\ & + \alpha_{18} \Delta NI_{t-1} * SIZE_t + \alpha_{19} D\Delta NI_{t-1} * \Delta NI_{t-1} * SIZE_t + \alpha_{20} LIT_t + \alpha_{21} D\Delta NI_{t-1} * LIT_t \\ & + \alpha_{22} \Delta NI_{t-1} * LIT_t + \alpha_{23} D\Delta NI_{t-1} * \Delta NI_{t-1} * LIT_t + \alpha_{24} IND_j + \alpha_{25} D\Delta NI_{t-1} * IND_j \\ & + \alpha_{26} \Delta NI_{t-1} * IND_j + \alpha_{27} D\Delta NI_{t-1} * \Delta NI_{t-1} * IND_j + \varepsilon \end{aligned} \tag{2}$$

where firm i subscripts are omitted.

Model (2) controls for firm characteristics that are used in prior research to proxy for the demand for conservatism. We include MB, the market value of equity divided by the book value of equity at the end of the fiscal year t, to control for the demand of conservatism arising from information asymmetry associated with a firm's growth option (LaFond and Watts 2008), and because studies document a negative association between conditional and unconditional conservatism (Givoly et al. 2007; Roychowdhury and Watts 2007). Firms with high levels of leverage tend to have greater bondholder and shareholder conflicts that, in turn, increase the contractual demand for conservatism (Ahmed et al. 2002; Zhang 2008). Hence, we include LEV, the sum of long-term debt and current liabilities deflated by total assets at the end of the fiscal year t.

LaFond and Watts (2008) argue that large firms produce more public information and have less information asymmetry, reducing the demand for conservative accounting. Hence, we control for firm size ($SIZE$) using the natural log of total assets at the end of the fiscal year t.[9] We also control for litigation risk because it can increase managers' incentive to recognize losses in a more timely manner than gains (Basu 1997; Watts 2003). Following Francis et al. (1994), we define an

[9] We mean-adjust all the continuous control variables in our regression models (i.e., MB, LEV, and $SIZE$) to reduce the problems with multicollinearity among the interaction terms (Neter et al. 1989; Aiken and West 1991). This procedure helps to make the magnitude of the coefficients more comparable to prior work.

indicator variable *LIT* that equals 1 if a firm operates in a litigious industry, and 0 otherwise.[10] Finally, Givoly et al. (2007) contend that the degree of conservatism varies across different industries. Hence, we include industry indicators (IND_j) following the industry classification in Frankel et al. (2002).[11]

H2 tests whether MW firms that remediate their MWs will exhibit different levels of conservatism from MW firms that continue to have these weaknesses. Following Ashbaugh-Skaife et al. (2008), we deem an unqualified SOX 404 opinion by the external auditor to objectively and unambiguously show that the firm has fully remediated its MWs. We test H2 using the sample of MW firms and replace *MW* with *FIXED* in model (2), where *FIXED* is an indicator variable that equals 1 if an MW firm receives an unqualified second SOX 404 report, and 0 otherwise. A negative (positive) coefficient on $D\Delta NI_{t-1} * \Delta NI_{t-1} * FIXED$ indicates that MW firms that remediate their weaknesses exhibit greater (lower) conservatism than MW firms that still have weaknesses.

Accrual-Based Loss Recognition (Ball and Shivakumar 2005)

Ball and Shivakumar (2005) develop a model to describe the differential timeliness of gain and loss recognition that relies on the correlation between accruals and contemporaneous cash flows.[12] In their regression, as shown in model (3) below, the authors predict a positive coefficient on $DCFO * CFO$ for accounting conservatism.

$$ACCRUAL = \gamma_0 + \gamma_1 DCFO + \gamma_2 CFO + \gamma_3 DCFO * CFO \tag{3}$$

where firm *i* and time *t* subscripts are omitted. *ACCRUAL* is net income before extraordinary items minus operating cash flows for firm *i* in fiscal year *t* deflated by beginning-of-year total assets. *CFO* is operating cash flows for firm *i* in fiscal year *t* deflated by beginning-of-year total assets. *DCFO* is an indicator variable that equals 1 if *CFO* is less than 0, and 0 otherwise. To examine H1, we develop model (4) as shown below. If MWs in internal controls result in lower (greater) accounting conservatism, then the accruals can less (more) effectively and timely reflect the future expectation of the negative change of cash flows; the coefficient on $DCFO * CFO * MW$ is predicted to be negative (positive).

$$\begin{aligned} ACCRUAL = {} & \gamma_0 + \gamma_1 DCFO + \gamma_2 CFO + \gamma_3 DCFO * CFO + \gamma_4 MW + \gamma_5 DCFO * MW \\ & + \gamma_6 CFO * MW + \gamma_7 DCFO * CFO * MW + \gamma_8 MB + \gamma_9 DCFO * MB \\ & + \gamma_{10} CFO * MB + \gamma_{11} DCFO * CFO * MB + \gamma_{12} LEV + \gamma_{13} DCFO * LEV \\ & + \gamma_{14} CFO * LEV + \gamma_{15} DCFO * CFO * LEV + \gamma_{16} SIZE + \gamma_{17} DCFO * SIZE \end{aligned}$$

---

10 Consistent with Francis et al. (1994), firms with primary SIC codes of 2833–2836 (biotechnology), 3570–3577 (computer equipment), 3600–3674 (electronics), 5200–5961 (retailing), and 7370–7374 (computer services) are considered to be operating in a litigious industry.

11 Following Frankel et al. (2002), industry membership is determined by the following SIC codes: agriculture (AGR, 0100–0999), mining and construction (MIN, 1000–1999, excluding 1300–1399), food (FOO, 2000–2111), textiles and printing/publishing (TEX, 2200–2799), chemicals (CHE, 2800–2824, 2840–2899), pharmaceuticals (PHA, 2830–2836), extractive (EXT, 2900–2999, 1300–1399), durable manufacturers (DUR, 3000–3999, excluding 3570–3579 and 3670–3679), transportation (TRA, 4000–4899), utilities (UTI, 4900–4999), retail (RET, 5000–5999), financial (FIN, 6000–6999), services (SER, 7000–8999, excluding 7370–7379) and computers (COM, 3570–3579, 3670–3679, 7370–7379).

12 The authors assert a positive but asymmetric correlation between accruals and contemporaneous cash flows. This positive correlation arises because cash flow revisions in the current period tend to be positively correlated with the current revisions for expected future cash flows. Timely recognition of unrealized gains and losses is based on expected, not realized, cash flows, and is therefore accomplished through accruals. Consequently, timely gain and loss recognition will produce a positive correlation between accruals and current period cash flows. This correlation is asymmetric because losses, under conservative reporting, are likely to be recognized on a timelier basis than gains.

$$+ \gamma_{18}CFO * SIZE + \gamma_{19}DCFO * CFO * SIZE + \gamma_{20}LIT + \gamma_{21}DCFO * LIT$$
$$+ \gamma_{22}CFO * LIT + \gamma_{23}DCFO * CFO * LIT + \gamma_{24}IND_j + \gamma_{25}DCFO * IND_j$$
$$+ \gamma_{26}CFO * IND_j + \gamma_{27}DCFO * CFO * IND_j + \varepsilon \quad (4)$$

where firm i and time t subscripts are omitted and all variables are as previously defined. We replace *MW* in model (4) with *FIXED* to test H2. If the remediation of MWs is associated with greater (lower) conservatism, then the coefficient on $DCFO * CFO * FIXED$ is predicted to be positive (negative).

Timeliness of Earnings to News (Basu 1997)

Our third measure of conservatism is a firm's timeliness of earnings to news (Basu 1997). As shown in model (5) below, the timeliness of earnings is inferred from the responsiveness of accounting income to the change in market values. Negative (positive) market-adjusted stock returns are used as proxies for bad (good) news. The asymmetric recognition of economic losses relative to gains is captured by a positive coefficient on $DR * R$.

$$NI = \beta_0 + \beta_1 DR + \beta_2 R + \beta_3 DR * R + \varepsilon \quad (5)$$

where firm i and time t subscripts are omitted. *NI* is the net income before extraordinary items for firm i in fiscal year t, deflated by the beginning-of-year market value, R is the market-adjusted stock return for firm i over the fiscal year t, and *DR* is an indicator variable that equals 1 if R is less than 0, and 0 otherwise.[13]

To test H1, we develop model (6). A negative (positive) coefficient on $DR * R * MW$ indicates that MW firms exhibit lower (higher) conservatism than firms without such weaknesses; that is, the MW firms have lower (higher) incremental timeliness of earnings to bad news than to good news. To test H2, we replace *MW* with *FIXED* in model (6). A positive (negative) coefficient on $DR * R * FIXED$ would suggest that MW firms that remediate their weaknesses exhibit greater (lower) conservatism than MW firms that fail to do so.

$$NI = \beta_0 + \beta_1 DR + \beta_2 R + \beta_3 DR * R + \beta_4 MW + \beta_5 DR * MW + \beta_6 R * MW + \beta_7 DR * R * MW$$
$$+ \beta_8 MB + \beta_9 DR * MB + \beta_{10} R * MB + \beta_{11} DR * R * MB + \beta_{12} LEV + \beta_{13} DR * LEV$$
$$+ \beta_{14} R * LEV + \beta_{15} DR * R * LEV + \beta_{16} SIZE + \beta_{17} DR * SIZE + \beta_{18} R * SIZE$$
$$+ \beta_{19} DR * R * SIZE + \beta_{20} LIT + \beta_{21} DR * LIT + \beta_{22} R * LIT + \beta_{23} DR * R * LIT + \beta_{24} IND_j$$
$$+ \beta_{25} DR * IND_j + \beta_{26} R * IND_j + \beta_{27} DR * R * IND_j + \varepsilon \quad (6)$$

where firm i and time t subscripts are omitted and all variables are as previously defined.

Givoly et al. (2007) point out the importance of controlling for the information or disclosure environment when using the Basu's (1997) timeliness measure. The control variables in our

[13] Dietrich et al. (2007) argue that the interpretation of model (5) is valid only when the market is efficient; in particular, if market returns cause earnings, and not the reverse. The authors also show that partitioning a regression sample with one of the regressors (R) may lead to biased inferences. However, Ryan (2006) notes that two well-known empirical results, the low R^2s observed in contemporaneous returns-earnings regressions (Collins et al. 1997; Ely and Waymire 1999; Francis and Schipper 1999) and a large literature showing that returns typically reflect information on a timelier basis than earnings, indicate that the concern mentioned by Dietrich et al. (2007) is likely to induce a very tiny bias in the estimation of conservatism. Ryan (2006) recommends measuring returns over the fiscal year to partially remove the impact of the annual earnings announcement over stock prices, which occurs approximately three months after closing. He also recommends the use of market-adjusted returns, defined as raw returns minus the value-weighted CRSP market return, to create the partitioning dummy variable *DR* in the Basu regression. Hence, to mitigate the concerns of Dietrich et al. (2007), we measure returns over the fiscal year and use market-adjusted returns instead of raw returns.

regressions partially address this issue.[14] Nonetheless, to further mitigate this concern, we control for other factors that are likely to proxy for the amount or quality of information disclosed by a firm—the annual frequency of management forecasts, the number of days from management forecast to fiscal year end date, audit quality (proxied by audit firm size), analyst following, and analyst forecast accuracy. The interpretation of our results for Basu's (1997) timeliness measure (untabulated) is not affected by the inclusion of these variables.

Sample Selection

Using the sample firms in Doyle et al. (2007a), we identify 1,098 firms that, under either SOX 302 or SOX 404, disclose at least one MW from January 2003 to November 2005.[15] We focus on firms that disclosed MWs because the reporting of MWs is mandatory, whereas the reporting of significant deficiencies and control deficiencies is not (Doyle et al. 2007a).[16] We deem a firm to have weak internal controls (i.e., low internal control quality) so long as it has at least one MW.

Table 1, Panel A summarizes the sample selection procedure for the descriptive statistics and for testing H1. We choose the fiscal years 2000 and 2001 to test H1 because these years just precede the enactment of SOX and, hence, avoid any confounding effects due to SOX. Our assumption is that MWs exist within the firm even before their disclosures from January 2003 to November 2005 (Doyle et al. 2007a). Based on the sample firms from Doyle et al. (2007a), we identify 1,230 (6,835) firm-year observations for the MW firms (control firms with no MW disclosures from January 2003 to November 2005) that have available data in Compustat and CRSP for fiscal years 2000 and 2001. Then we remove 84 (434) outlier observations for all the continuous variables at the top and bottom 0.5 percent levels, leaving us with 1,146 (6,401) firm-year observations for the MW firms (control firms) sample.[17] In total, we have 7,547 firm-year observations for the descriptive statistics and H1 tests.

Table 1, Panel B summarizes the sample selection procedure for testing H2, our remediation test. We focus on the MW sample because remediation is not applicable to the control sample. Based on an initial 1,098 MW firms, we remove 473 firms that have no second SOX 404 opinions; that is, we are not sure whether the MWs are remediated. These firms are either non-accelerated filers that are not required to file SOX 404 opinions (289 firms) or they have terminated registra-

[14] First, Givoly et al. (2007) contend that the timeliness measure is understated for firms where the information environment is characterized by a smooth and frequent arrival of news relative to firms for which the news arrives less frequently. The information environment of larger firms is more likely to conform to the former characterization. In this aspect, controlling for firm size addresses this concern. Second, Givoly et al. (2007) caution that if a highly litigious environment prompts management to preempt bad news by disclosing it early, the Basu's timeliness measure will under-estimate the degree of conservatism. Our inclusion of *LIT* as control in the regression mitigates this concern. Third, Givoly et al. (2007) point out that Basu's (1997) measure of conservatism is likely to be negatively related to unconditional conservatism. Hence, we include market-to-book value for control of unconditional conservatism. Fourth, Givoly et al. (2007) highlight that the information environment can vary across industries, which will affect the timeliness measure. We address this concern through the inclusion of industry-dummy variables as controls.

[15] We thank Jeffrey Doyle, Weili Ge, and Sarah McVay for sharing the data. The data can be found at: http://faculty.washington.edu/geweili/ICdata.html.

[16] Although both MWs and significant deficiencies are deficiencies in the design or operation of internal controls, significant deficiencies are less severe and are not required to be publicly disclosed under SOX 302 (SEC 2004). Hence, the disclosure of significant deficiencies is clearly voluntary. On the other hand, under SOX 302, if management identifies an MW in their controls, then they are precluded from reporting that the controls are effective and must disclose the identified MW. Hence, the disclosure of MWs is effectively mandatory. According to Doyle et al. (2007b), there is some ambiguity regarding whether SOX 302 certifications require the public disclosure of MWs and whether some firms might interpret the MW disclosure requirement under SOX 302 as voluntary. The authors' conclusion, from reading most of the SEC guidance, is that most firms are treating the disclosure as mandatory.

[17] Because we use the same sample for the descriptive statistics and for all of the H1 tests (based on the three different conservatism measures), an outlier observation is removed from all the H1 tests, rather than from just the test to which it applies.

TABLE 1
Sample Selection Procedure

Panel A: Sample Selection for H1

	Observations
MW Sample	
Observations with available Compustat and CRSP data for 2000 and 2001	1,230
Less: outlier observations for all continuous variables at the top and bottom 0.5% levels	(84)
Total MW firm observations	1,146
Control Sample	
Observations with available Compustat and CRSP data for 2000 and 2001	6,835
Less: outlier observations for all continuous variables at the top and bottom 0.5% levels	(434)
Total control firm observations	6,401
Total firm-year observations used in descriptive statistics and testing H1	7,547

Panel B: Sample Selection for H2

	Observations
Identified MW firms from 2003 to 2005	1,098
less: firms without SOX 404 auditor opinions due to non-accelerated filers status	(289)
less: firms without SOX 404 auditor opinions due to termination of securities registration	(184)
less: firms with missing data in Compustat and CRSP	(109)
Total firm-year observations used in testing H2	516*

* Among the 516 firms with a second SOX 404 opinion, 366 received an unqualified auditor report and 150 received an adverse report.

tion of their securities (184 firms). We also remove 109 firms with missing data in Compustat and CRSP. This procedure yields a final sample of 516 firms for the H2 tests, with 366 unqualified opinions and 150 adverse opinions.

IV. RESULTS

Descriptive Statistics

Table 2, Panel A presents the descriptive statistics for the sample used to test H1, which is based on the firm-year observations for the fiscal years 2000 and 2001. Consistent with Doyle et al. (2007b) and Ashbaugh-Skaife et al. (2007), the table shows that MW firms are smaller, financially weaker, and more leveraged than the control firms. The MW firms are also more likely to operate in a litigious industry, and have more negative change in net income in year $t-1$ and more negative accruals than the control firms. We do not observe any differences in the change in net income in year t, stock returns, cash flow from operations, stock return, and market-to-book ratio between the MW and control firms. Table 2, Panel B presents both Pearson and Spearman correlation statistics; we do not find any unusual correlations among the independent variables in our regressions that warrant concern.

Multivariate Regression Results

Regression Results for H1

Table 3 presents the regression results of H1 based on the persistence of earnings changes measure. We first run a baseline regression model without the *MW* and its related terms. This allows us to assess the incremental explanatory power of internal control quality on conservatism once the *MW* terms are included into the estimations. The baseline regression results, presented in Column 1, show that the coefficient on $D\Delta NI_{t-1} * \Delta NI_{t-1}$ is -0.368 ($p < 0.01$), confirming that financial reporting is conservative in general. Column 2 shows that, after including the *MW* and the related terms, the adjusted R^2 of the regression increases from 24.0 percent to 24.6 percent. This result suggests that internal control quality provides incremental explanatory power to conservatism in addition to factors affecting the demand for conservatism. Column 2 also reveals that the coefficient on $D\Delta NI_{t-1} * \Delta NI_{t-1} * MW$ is 0.217 ($p = 0.01$), which implies that the MW firms have a reduced tendency to reverse negative earnings changes in the following period (i.e., are less conservative) relative to the control firms. This result is consistent with a positive relation between internal control quality and conservatism.

We repeat the above analyses using the accrual-based loss recognition measure and present the results in Table 4. Column 1 shows that the coefficient on $DCFO * CFO$ is 0.669 ($p < 0.01$), indicating the presence of reporting conservatism. Column 2 shows that the addition of *MW* and its related terms to the baseline regression increases the adjusted R^2 from 16.7 percent to 17.0 percent, suggesting that internal control quality provides incremental explanatory power to a firm's level of conservatism. The coefficient on $DCFO * CFO * MW$ is -0.149 ($p = 0.03$), indicating that the MW firms accrue less unrealized losses in the cash-loss year; this result provides further support that the MW firms exhibit lower conservatism than the control firms.

Finally, Table 5 presents the results of the same analyses using the timeliness of earnings to news measure. Column 1 shows that the coefficient on $DR * R$ is 0.402 ($p < 0.01$), consistent with the presence of conservative reporting. When we add the *MW* and its related terms to the baseline regression, the adjusted R^2 of the model increases from 28.0 percent to 29.0 percent. Column 2 further reveals that the coefficient on $DR * R * MW$ is -0.072 ($p = 0.06$), which implies that the earnings of the MW firms reflect unexpected losses in a less timely manner than the control firms (i.e., the MW firms are less conservative). Again, this finding is consistent with the above results

TABLE 2
Descriptive Statistics and Correlation Analysis

Panel A: Distributional Properties of Variables

	Mean	Median	Std. Dev.	Q1	Q3
MW Firms (n = 1,146 observations in 2000 and 2001)					
ΔNI_t	−0.02	0.00	0.16	−0.06	0.03
ΔNI_{t-1}	−0.04**	0.01	0.12	−0.18	0.04
ACCRUAL	−0.08**	−0.06*	0.14	−0.13	−0.01
CFO	0.05	0.06**	0.16	−0.01	0.13
NI	−0.06*	0.02**	0.24	−0.07	0.07
R	0.18	0.03*	0.84	−0.37	0.47
MB	2.69	1.64	3.69	0.94	3.06
LEV	0.19**	0.13**	0.20	0.01	0.32
SIZE-ACTUAL	1811.17***	274.01*	6882.37	87.76	915.60
LIT	0.38***	0.00***	0.49	0.00	1.00
Control Firms (n = 6,401 observations in 2000 and 2001)					
ΔNI_t	−0.02	0.00	0.14	−0.04	0.03
ΔNI_{t-1}	−0.03**	0.00	0.36	−0.14	0.04
ACCRUAL	−0.07**	−0.05*	0.13	−0.11	−0.01
CFO	0.05	0.07**	0.17	0.00	0.13
NI	−0.05*	0.03**	0.23	−0.05	0.08
R	0.19	0.06*	0.78	−0.31	0.48
MB	2.68	1.66	3.90	0.92	3.06
LEV	0.18**	0.12**	0.19	0.00	0.30
SIZE-ACTUAL	3073.94***	301.64*	10489.78	74.01	1457.23
LIT	0.30***	0.00***	0.46	0.00	1.00

Panel B: Pearson Correlations (top) and Spearman Correlations (bottom)

	ΔNI_t	ΔNI_{t-1}	*ACCRUAL*	*CFO*	*NI*	*R*	*MB*	*LEV*	*SIZE*	*LIT*
ΔNI_t	1.00	−0.18†	0.48†	0.31†	0.44†	0.32†	0.04†	0.04†	0.05†	−0.12†
ΔNI_{t-1}	−0.06†	1.00	0.18†	0.24†	0.25†	0.05†	0.03†	0.04†	0.08†	−0.06†

(continued on next page)

Panel B: Pearson Correlations (top) and Spearman Correlations (bottom)

	ΔNI_t	ΔNI_{t-1}	*ACCRUAL*	*CFO*	*NI*	*R*	*MB*	*LEV*	*SIZE*	*LIT*
ACCRUAL	0.31†	0.13†	1.00	−0.07†	0.40†	0.11†	−0.05†	0.04†	0.09†	−0.16†
CFO	0.30†	0.27†	−0.28†	1.00	0.48†	0.24†	−0.03†	0.10†	0.28†	−0.18†
NI	0.41†	0.24†	0.46†	0.43†	1.00	0.27†	0.08†	0.07†	0.24†	−0.19†
R	0.22†	0.03†	0.11†	0.13†	0.26†	1.00	0.23†	0.01	0.02	−0.07†
MB	0.17†	0.15†	0.00	0.17†	−0.05†	0.30†	1.00	−0.07†	0.03†	0.16†
LEV	0.04†	0.01	0.01	0.09†	0.19†	0.08†	−0.11†	1.00	0.33†	−0.27†
SIZE	0.06†	0.08†	0.07†	0.25†	0.23†	0.12†	0.14†	0.42†	1.00	−0.18†
LIT	−0.09†	−0.03†	−0.14†	−0.13†	−0.27†	−0.15†	0.17†	−0.30†	−0.19†	1.00

*, **, *** Denote significant differences between the MW sample and the control sample at the 10 percent, 5 percent, and 1 percent levels, respectively, based on two-tailed tests. Differences in means (medians) are assessed using a t-test (Wilcoxon rank sum test).

† Indicates significance at the 5 percent level, two-tailed.

Panel A shows the distribution of the MW firms and the control firms with no MWs.

Panel B shows the correlations among the key variables (identified in Panel A) used in the empirical analyses. Pearson and Spearman correlations are found, respectively, above and below the diagonal.

Variable Definitions:

ΔNI_t = change in net income before extraordinary items for firm *i* in fiscal year *t* deflated by beginning-of-year total assets;

ΔNI_{t-1} = change in net income before extraordinary items for firm *i* in fiscal year *t*−1 deflated by beginning-of-year total assets;

ACCRUAL = net income before extraordinary items minus operating cash flows (Data308) for firm *i* in fiscal year *t* deflated by beginning-of-year total assets;

CFO = operating cash flows at the end of the fiscal year deflated by beginning-of-year total assets;

NI = net income before extraordinary items (Data123) for firm *i* in fiscal year *t* deflated by the beginning-of-year market value of equity (Data199 * Data25);

R = market-adjusted stock return for firm *i* over the fiscal year *t*;

MB = market value of equity (Data199 * Data25) divided by the book value of equity (Data60) at the end of the fiscal year *t*;

LEV = firm's leverage, measured by the sum of long-term debt (Data9) and debt in current liabilities (Data34) deflated by market value of equity at the end of the fiscal year *t*;

SIZE = natural log of the total assets at the end of the fiscal year *t*. We use *SIZE* as the control variable in the regressions but for the descriptive statistics, we report the untransformed value of total assets (*SIZE-ACTUAL*); and

LIT = indicator variable that equals 1 if a firm operates in a litigious industry, and 0 otherwise. Consistent with Francis et al. (1994), firms with primary SIC codes of 2833–2836 (biotechnology), 3570–3577 (computer equipment), 3600–3674 (electronics), 5200–5961 (retailing), and 7370–7374 (computer services) are considered to be operating in a litigious industry.

TABLE 3

Regression Results of H1 using the Persistence of Earnings Changes Measure

Independent Variables	Pred. Sign	Baseline Regression (1) Coeff.	p-value	Full Regression (2) Coeff.	p-value
Intercept	?	−0.042***	<0.01	−0.043***	<0.01
$D\Delta NI_{t-1}$	?	−0.026***	<0.01	−0.025***	<0.01
ΔNI_{t-1}	?	−0.145***	0.01	−0.099*	<0.01
$D\Delta NI_{t-1} * \Delta NI_{t-1}$	−	−0.368***	<0.01	−0.412***	<0.01
MW	?			0.007	0.26
$D\Delta NI_{t-1} * MW$	?			−0.009	0.36
$\Delta NI_{t-1} * MW$	?			−0.246***	<0.01
$D\Delta NI_{t-1} * \Delta NI_{t-1} * MW$	?			0.217***	0.01
MB_t	?	0.005***	<0.01	0.005***	<0.01
$D\Delta NI_{t-1} * MB_t$	?	−0.008***	<0.01	−0.008***	<0.01
$\Delta NI_{t-1} * MB_t$	?	−0.017***	<0.01	−0.016***	<0.01
$D\Delta NI_{t-1} * \Delta NI_{t-1} * MB_t$	?	0.024***	<0.01	0.023***	<0.01
LEV_t	?	0.086***	<0.01	0.085***	<0.01
$D\Delta NI_{t-1} * LEV_t$	?	0.006	0.47	0.007	0.39
$\Delta NI_{t-1} * LEV_t$	?	0.395***	<0.01	0.406***	<0.01
$D\Delta NI_{t-1} * \Delta NI_{t-1} * LEV_t$	−	−0.374***	<0.01	−0.385***	<0.01
$SIZE_t$	?	0.005***	<0.01	0.004***	<0.01
$D\Delta NI_{t-1} * SIZE_t$	?	0.002	0.31	0.002	0.27
$\Delta NI_{t-1} * SIZE_t$	?	−0.096***	<0.01	−0.092***	<0.01
$D\Delta NI_{t-1}\Delta NI_{t-1} * SIZE_t$	+	0.149***	<0.01	0.144***	<0.01
LIT_t	?	−0.010	0.14	−0.010	0.16
$D\Delta NI_{t-1} * LIT_t$	?	−0.045***	<0.01	−0.046***	<0.01
$\Delta NI_{t-1} * LIT_t$	?	−0.177**	0.02	−0.187***	0.01
$D\Delta NI_{t-1} * \Delta NI_{t-1} * LIT_t$	−	0.137	0.15	0.147	0.13
Industry Effects		Not reported		Not reported	
Adjusted R^2		24.0%		24.6%	
Number of observations		7,547		7,547	

*, **, *** Denote significance at the 10 percent, 5 percent, and 1 percent levels, respectively, based on two-tailed tests.

This table reports the regression results of H1 using the persistence of earnings changes measure of conservatism in Basu (1997). Column 1 (Column 2) shows the results without (with) the *MW* and its related terms. The dependent variable is ΔNI_t, which is the change in net income before extraordinary items for firm *i* in fiscal year *t* deflated by beginning-of-year total assets. *MW* is an indicator variable that equals 1 if the firm has at least one MW in internal controls, and 0 otherwise. The other variables are as previously defined. A positive (negative) coefficient on $D\Delta NI_{t-1} * \Delta NI_{t-1} * MW$ indicates that MW firms exhibit lower (greater) conservatism than the control firms.

of a positive relation between internal control quality and conservatism.[18]

[18] Prior research shows that conservatism is associated with governance factors such as managerial ownership (LaFond and Roychowdhury 2008) and board of director characteristics (Ahmed and Duellman 2007). Hence, we also examine whether internal control quality has an incremental effect on conservatism after controlling for these factors. We repeat the H1 tests in Tables 3–5 by further adding CEO share ownership, institutional ownership, the duality of the CEO and

TABLE 4
Regression Results of H1 Using the Accrual-Based Loss Recognition Measure

Independent Variables	Pred. Sign	Baseline Regression (1)		Full Regression (2)	
		Coeff.	p-value	Coeff.	p-value
Intercept	?	−0.020***	0.01	−0.014*	0.09
DCFO	?	0.002	0.86	−0.026*	0.10
CFO	−	−0.502***	<0.01	−0.534***	<0.01
*DCFO * CFO*	+	0.669***	<0.01	0.643***	<0.01
MW	?			−0.013*	0.07
*DCFO * MW*	?			−0.006	0.64
*CFO * MW*	?			0.048	0.34
*DCFO * CFO * MW*	?			−0.149**	0.03
MB	?	0.003***	<0.01	0.003***	<0.01
*DCFO * MB*	?	−0.004***	<0.01	−0.004***	0.01
*CFO * MB*	?	0.000	1.00	0.000	0.95
*DCFO * CFO * MB*	?	0.000	0.98	0.000	0.93
LEV	?	−0.023	0.11	−0.021	0.15
*DCFO * LEV*	?	0.004	0.88	−0.006	0.82
*CFO * LEV*	?	−0.232**	0.03	−0.241**	0.03
*DCFO * CFO * LEV*	+	0.455***	<0.01	0.438***	<0.01
SIZE	?	0.003**	0.03	0.003**	0.03
*DCFO * SIZE*	?	0.000	0.97	0.000	0.97
*CFO * SIZE*	?	0.006	0.56	0.005	0.59
*DCFO * CFO * SIZE*	−	−0.070***	<0.01	−0.070***	<0.01
LIT	?	−0.009	0.40	−0.008	0.46
*DCFO * LIT*	?	−0.024	0.16	−0.023	0.17
*CFO * LIT*	?	0.046	0.58	0.041	0.62
*DCFO * CFO * LIT*	+	0.083	0.42	0.104	0.31
Industry Effects		Not reported		Not reported	
Adjusted R^2		16.7%		17.0%	
Number of observations		7,547		7,547	

*, **, *** Denote significance at the 10 percent, 5 percent, and 1 percent levels, respectively, based on two-tailed tests.
This table reports the regression results of H1 using the accrual-based loss recognition measure of conservatism in Ball and Shivakumar (2005). Column 1 (Column 2) shows the results without (with) the *MW* and its related terms. The dependent variable is *ACCRUAL*, which is net income before extraordinary items minus operating cash flows for firm *i* in fiscal year *t* deflated by beginning-of-year total assets. The other variables are as previously defined. A negative (positive) coefficient on *DCFO * CFO * MW* indicates that MW firms exhibit lower (greater) conservatism than the control firms.

Our findings on the control variables are generally consistent with prior research. For instance, the coefficients on $D\Delta NI_{t-1} * \Delta NI_{t-1} * MB$ and $DR * R * MB$ confirm the negative relation

Chairman positions, board size, board independence, the share ownership of the independent directors, and the average outside directorships of the independent directors. The sample size for these tests is 1,920, which is much smaller than that used in the analyses in Tables 3–5. This is because we obtain the data on managerial ownership from the ExecuComp database and the governance characteristics from the RiskMetrics database. Both databases provide information on S&P 1500 companies only. Our inferences on the results for H1 (untabulated) remain unchanged. Specifically, the coefficients on $D\Delta NI_{t-1} * \Delta NI_{t-1} * MW$, $DCFO * CFO * MW$, and $DR * R * MW$ are 0.423 (p = 0.06), −0.510 (p = 0.03), and −0.089 (p = 0.07), respectively.

TABLE 5

Regression Results of H1 using the Timeliness of Earnings to News Measure

Independent Variables	Pred. Sign	Baseline Regression (1) Coeff.	p-value	Full Regression (2) Coeff.	p-value
Intercept	?	−0.011*	0.07	−0.009	0.13
DR	?	0.049***	<0.01	0.052***	<0.01
R	?	0.011**	0.05	0.013**	0.03
*DR * R*	+	0.402***	<0.01	0.414***	<0.01
MW	?			−0.010	0.39
*DR * MW*	?			−0.017	0.41
*R * MW*	?			−0.011	0.29
*DR * R * MW*	?			−0.072*	0.06
MB	?	0.001	0.37	0.001	0.37
*DR * MB*	?	−0.004**	0.05	−0.004*	0.06
*R * MB*	?	−0.001	0.23	−0.001	0.22
*DR * R * MB*	?	−0.020***	<0.01	−0.020***	<0.01
LEV	?	−0.037	0.12	−0.033	0.15
*DR * LEV*	?	0.018	0.67	0.017	0.69
*R * LEV*	?	0.032	0.19	0.030	0.22
*DR * R * LEV*	+	0.372***	<0.01	0.378***	<0.01
SIZE	?	0.010***	<0.01	0.009***	<0.01
*DR * SIZE*	?	−0.005	0.21	−0.004	0.23
*R * SIZE*	?	−0.001	0.84	−0.001	0.79
*DR * R * SIZE*	−	−0.077***	<0.01	−0.076***	<0.01
LIT	?	−0.063***	<0.01	−0.062***	<0.01
*DR * LIT*	?	0.011	0.57	0.012	0.55
*R * LIT*	?	0.013	0.22	0.014	0.19
*DR * R * LIT*	+	−0.034	0.32	−0.033	0.34
Industry Effects		Not reported		Not reported	
Adjusted R^2		28.0%		29.0%	
Number of observations		7,547		7,547	

*, **, *** Denote significance at the 10 percent, 5 percent, and 1 percent levels, respectively, based on two-tailed tests.
This table reports the regression results of H1 using the timeliness of earnings to news measure of conservatism in Basu (1997). Column 1 (Column 2) shows the results without (with) the *MW* and its related terms. The dependent variable is *NI*, which is the net income before extraordinary items for firm *i* in fiscal year *t* deflated by the beginning-of-year market value of equity. The other variables are as previously defined. A negative (positive) coefficient on *DR * R * MW* indicates that MW firms exhibit lower (greater) conservatism than the control firms.

between conservatism and unconditional conservatism documented in other studies.[19] The coefficients on $D\Delta NI_{t-1} * \Delta NI_{t-1} * LEV$, $DCFO * CFO * LEV$, and $DR * R * LEV$ are also consistent with prior research that shows that more highly leveraged firms exhibit more conservatism. The

[19] We also use another measure of unconditional conservatism devised by Penman and Zhang (2002), which is the amount of a firm's "hidden" reserves estimated using the sum of the value of the LIFO reserve, the research and development reserve, and the advertising reserve. Using this alternative measure, we continue to find a significant and positive relation between internal control quality and conservatism, and a significant and negative relation between unconditional conservatism and our conservatism measures.

coefficients on $D\Delta NI_{t-1} * \Delta NI_{t-1} * SIZE$, $DCFO * CFO * SIZE$, and $DR * R * SIZE$ suggest that larger firms have lower conservatism, consistent with the aggregation effect discussed in Givoly et al. (2007) and with the information asymmetry hypothesis discussed in LaFond and Watts (2008). Finally, similar to LaFond and Roychowdhury (2008), we do not find a significant relation between the level of litigation risks and conservatism.

Regression Results for H2

Table 6 reports the regression results for testing H2 using the persistence of earnings changes measure (Column 1), accrual-based loss recognition measure (Column 2), and timeliness of earnings to news measure (Column 3). For brevity, we present only the coefficients of the three-way interaction terms for the control variables. Column 1 reveals that the coefficient on $D\Delta NI_{t-1} * \Delta NI_{t-1} * FIXED$ is −0.622 (p = 0.03). This result shows that MW firms that remediate

TABLE 6

Regression Results of H2

		Persistence of Earnings Changes $A = D\Delta NI_{t-1}$ $B = \Delta NI_{t-1}$ (1)		Accrual-Based Loss Recognition $A = DCFO$ $B = CFO$ (2)		Timeliness of Earnings to News $A = DR$ $B = R$ (3)	
	Pred. Sign	Coeff.	p-value	Coeff.	p-value	Coeff.	p-value
Intercept	?	−0.004	0.85	−0.020	0.28	−0.017	0.56
A	?	−0.004	0.91	−0.111***	<0.01	−0.003	0.95
B	?	−0.553***	<0.01	−0.803***	<0.01	0.046	0.39
*A * B*	–, +, +	0.160	0.54	0.436*	0.10	0.281***	<0.01
FIXED	?	0.014	0.59	0.016	0.40	0.039	0.19
*A * FIXED*	?	−0.018	0.63	0.097***	<0.01	0.069	0.11
*B * FIXED*	?	0.240	0.29	0.205	0.17	−0.002	0.97
*A * B * FIXED*	?	−0.622**	0.03	0.748***	<0.01	0.186*	0.07
*A * B * MB*	?	0.050	0.15	−0.041***	0.01	0.013	0.16
*A * B * LEV*	–, +, +	2.082**	0.05	0.397	0.38	0.391*	0.07
*A * B * SIZE*	+, –, –	0.021	0.85	−0.201***	0.01	0.036	0.24
*A * B * LIT*	–, +, +	−0.700**	0.05	−0.393	0.17	−0.241***	0.04
Industry Effects		Not reported		Not reported		Not reported	
Adjusted R^2		30.7%		43.1%		33.6%	
Number of observations		516		516		516	

*, **, *** Denote significance at the 10 percent, 5 percent, and 1 percent levels, respectively, based on two-tailed tests.

This table reports the regression results of H2 using the persistence of earnings changes measure (Column 1), accrual-based loss recognition measure (Column 2), and timeliness of earnings to news measure (Column 3). The dependent variables for Columns 1, 2, and 3 are ΔNI_t, *ACCRUAL*, and *NI*, respectively. *FIXED* is an indicator variable that equals 1 if an MW firm receives an unqualified second SOX 404 report, and 0 otherwise. All the other variables are as previously defined. For brevity, we present only the coefficients of the three-way interaction terms of the control variables. A negative (positive, positive) coefficient on $D\Delta NI_{t-1} * \Delta NI_{t-1} * FIXED$ ($DCFO * CFO * FIXED$, $DR * R * FIXED$) indicates that MW firms that remediate their weaknesses at the time of the second SOX 404 report exhibit greater conservatism than MW firms that fail to do so. A positive (negative, negative) coefficient on $D\Delta NI_{t-1} * \Delta NI_{t-1} * FIXED$ ($DCFO * CFO * FIXED$, $DR * R * FIXED$) indicates that MW firms that remediate their weaknesses at the time of the second SOX 404 report exhibit lower conservatism than MW firms that fail to do so.

their weaknesses at the time of the second SOX 404 report exhibit greater conservatism than MW firms that continue to have weaknesses at the time of the report. The results using the other two measures of conservatism are consistent with this finding. Specifically, the coefficient on $DCFO * CFO * FIXED$ in Column 2 is 0.748 ($p < 0.01$), and the coefficient on $DR * R * FIXED$ in Column 3 is 0.186 ($p = 0.07$). Hence, the regression results on H2 using all three measures of conservatism strengthen our earlier finding of a positive relation between internal control quality and conservatism.

In summary, the empirical findings using all three measures of conservatism show that internal control quality provides incremental explanatory power to a firm's level of conservatism. We find consistent evidence of a positive relation between internal control quality and conservatism. Specifically, firms with MWs exhibit lower conservatism than firms without such weaknesses, and MW firms that subsequently remediate their weaknesses exhibit greater conservatism than MW firms that fail to do so. Taken together, our results show that effective internal controls act as a mechanism that facilitates conservatism, consistent with the agency framework of positive accounting theory that recognizes conservatism as a governance mechanism (Watts 2003; Ball and Shivakumar 2005). In the next section, we describe additional analyses used to test the robustness of these results.

V. ADDITIONAL ANALYSES

Self-Selection of Internal Control Quality (H1)

Prior research suggests that MW firms are likely to be systematically different from other firms (Ashbaugh-Skaife et al. 2007; Doyle et al. 2007b). This could result in potential self-selection bias in our observed samples for our H1 tests. Although the tests on the change in the status of internal controls mitigate this concern, we further econometrically control for this bias using three methods. First, we use Heckman's (1979) two-stage procedure, in which the first stage comprises a probit regression of *MW* on its determinants. These determinants largely follow those of Doyle et al. (2007b) and Ashbaugh-Skaife et al. (2007), and are detailed in Table 7.[20] From this regression that identifies the likelihood of a firm being "selected" as an MW firm, we calculate the inverse Mills ratio, *LAMBDA* (see Heckman 1979; Leuz and Verrecchia 2000), and include it in our main regressions to control for the likelihood of self-selection into the MW group.

Second, we create a matched sample based on the predicted probabilities from the first-stage probit regression. This method, known as propensity score-matching (LaLonde 1986), creates a non-MW control sample with the same predicted probabilities of having an MW (i.e., it considers all predictive variables together to estimate the same propensity of having an MW). The propensity score-matching produces a matched sample of 850 control firm observations, resulting in a combined sample of 1,700 firm observations.[21] Third, we include the determinants of MWs, as well as their two-way and three-way interactions with the parameters of the three conservatism measures, as additional control variables in the main regressions presented earlier. This method controls for the possibility that the factors that produce the MWs could also be driving observed differences in conservatism.

Table 7 presents the results of the above three tests for the persistence of earnings changes measure. For brevity, we present only the results of the three-way interaction terms for the control variables. Column 1 shows the second-stage regression results of the Heckman two-stage proce-

---

[20] The first-stage regression results (untabulated) show that financial health (*LOSS*), rapid growth (*EXTR_SALES*), audit quality (*BIGN*), the occurrence of a restatement (*RESTATE*), and the occurrence of an auditor change (*AUDCHANGE*) are significantly associated with the presence of MWs. Details of the results are available from the authors upon request.

[21] The search algorithm finds the propensity scores closest to our MW firm observations until the first decimal place. The algorithm is available upon request.

TABLE 7

Regression Results of H1 Using the Persistence of Earnings Changes Measure and after Controlling for Self-Selection of Internal Control Quality

		Heckman Second-Stage Regression (1)		Propensity Score-Matching (2)		Controlling for Determinants of MWs (3)	
	Pred. Sign	Coeff.	p-value	Coeff.	p-value	Coeff.	p-value
Intercept	?	−0.086***	<0.01	−0.047***	<0.01	−0.033***	0.01
$D\Delta NI_{t-1}$	?	−0.022***	0.01	−0.038*	0.08	−0.037*	0.06
ΔNI_{t-1}	?	−0.014	0.82	−0.286***	<0.01	0.128	0.37
$D\Delta NI_{t-1} * \Delta NI_{t-1}$	−	−0.718***	<0.01	0.143	0.36	−0.384*	0.08
MW	?	0.016**	0.02	0.014	0.25	0.011*	0.10
$D\Delta NI_{t-1} * MW$	?	−0.016	0.15	0.003	0.86	−0.008	0.47
$\Delta NI_{t-1} * MW$	?	−0.422***	<0.01	−0.141*	0.06	−0.346***	<0.01
$D\Delta NI_{t-1} * \Delta NI_{t-1} * MW$	?	0.503***	<0.01	0.276***	0.01	0.445***	<0.01
$D\Delta NI_{t-1} * \Delta NI_{t-1} * MB_t$	?	0.023***	0.01	−0.024**	0.02	0.012	0.17
$D\Delta NI_{t-1} * \Delta NI_{t-1} * LEV_t$	−	−0.553***	<0.01	−0.839***	<0.01	−0.444***	<0.01
$D\Delta NI_{t-1} * \Delta NI_{t-1} * SIZE_t$	+	0.045**	0.05	0.208***	<0.01	0.014	0.59
$D\Delta NI_{t-1} * \Delta NI_{t-1} * LIT_t$	−	0.154	0.16	−0.761***	<0.01	0.088	0.37
$LAMBDA$	?	0.025***	<0.01				
$D\Delta NI_{t-1} * \Delta NI_{t-1} * AGE_t$	?					0.035	0.47
$D\Delta NI_{t-1} * \Delta NI_{t-1} * LOSS_t$	?					−0.485***	<0.01
$D\Delta NI_{t-1} * \Delta NI_{t-1} * SEGMENT_t$	?					−0.026	0.21
$D\Delta NI_{t-1} * \Delta NI_{t-1} * FOREIGN_t$	?					−0.223***	0.01
$D\Delta NI_{t-1} * \Delta NI_{t-1} * MA_t$	?					0.549*	0.10
$D\Delta NI_{t-1} * \Delta NI_{t-1} * RESTR_t$	?					2.079***	<0.01
$D\Delta NI_{t-1} * \Delta NI_{t-1} * EXTR\_SALES_t$	?					0.283***	<0.01
$D\Delta NI_{t-1} * \Delta NI_{t-1} * BIGN_t$	?					0.121	0.30
$D\Delta NI_{t-1} * \Delta NI_{t-1} * RESTATE_t$	?					0.824***	<0.01
$D\Delta NI_{t-1} * \Delta NI_{t-1} * AUDCHANGE_t$	?					−0.143*	0.08

(continued on next page)

TABLE 7 (continued)

	Pred. Sign	Heckman Second-Stage Regression (1)		Propensity Score-Matching (2)		Controlling for Determinants of MWs (3)	
		Coeff.	p-value	Coeff.	p-value	Coeff.	p-value
Industry Effects		Not reported		Not reported		Not reported	
Adjusted R^2		27.3%		17.2%		32.7%	
Number of observations		5,439		1,700		5,439	

*, **, *** Denote significance at the 10 percent, 5 percent, and 1 percent levels, respectively, based on two-tailed tests.

This table reports the results of the regressions that control for self-selection of internal control quality using the persistence of earnings changes measure. The dependent variable is ΔNI_t, which is the change in net income before extraordinary items for firm i in fiscal year t deflated by beginning-of-year total assets. For brevity, other than *MW*, we present only the coefficients of the three-way interaction terms of the control variables. Column 1 presents the second stage results of the Heckman (1979) two-stage procedure, after controlling for *LAMBDA*. *LAMBDA* is the inverse Mill ratio calculated from the first-stage regression in which we regress *MW* on its determinants. Column 2 presents the results of the propensity-score-matching approach. Column 3 presents the results after controlling for the determinants of MWs. We use the following determinants of MWs. Firm size is measured by the log of market capitalization, which is share price (Data25) multiplied by number of shares outstanding (Data199). Because we have already controlled for *SIZE*, we do not add market capitalization as further control in Column 3. The other variables are defined as follows or as previously defined.

Variable Definitions:

AGE = firm age, the number of years since the firm appears in CRSP database;
LOSS = indicator variable that equals 1 if net income before extraordinary items (Data123) in years t and t–1 sum to less than 0, and 0 otherwise;
SEGMENT = log of the sum of the number of operating and geographic segments reported by the Compustat Segments database for the firm in year t;
FOREIGN = indicator variable that equals 1 if the firm has a non-zero foreign currency translation (Data150), and 0 otherwise;
MA = indicator variable that equals 1 if the firm has a non-zero merger and acquisition activity (Data360);
RESTR = aggregate restructuring charges [Data376 × (−1)] in years t and t–1, scaled by the firm's market capitalization at the end of the fiscal year;
EXTR_SALES = indicator variable that equals 1 if year-over-year industry-adjusted sales growth (Data12) falls into the top quintile, and 0 otherwise;
BIGN = indicator variable that equals 1 if the firm engaged one of the largest four audit firms, and 0 otherwise. Largest four audit firms include PWC, Deloitte & Touche, Ernst & Young, and KPMG;
RESTATE = indicator variable that equals 1 if the firm had a restatement in the 12-month period before the disclosure of MWs, and 0 otherwise; and
AUDCHANGE = indicator variable that equals 1 if the firm changed auditor during the 12-month period before the disclosure of MWs, and 0 otherwise.

dure after controlling for *LAMBDA*. Column 2 shows the propensity score-matching results and Column 3 shows the results after controlling for the determinants of MWs. In all three regressions, we find that the coefficient on $D\Delta NI_{t-1} * \Delta NI_{t-1} * MW$ remains positive and significant, confirming our earlier findings in Table 3. When we repeat these analyses for the other two measures of conservatism, the results for all the three tests (untabulated) show that the coefficients on $DCFO * CFO * MW$ and $DR * R * MW$ remain negative and significant ($p < 0.10$).[22] Taken together, we interpret these findings as suggesting that self-selection issues are not likely to drive the results of H1 that we obtained earlier.

Self-Selection of Improvement in Internal Control Quality (H2)

It is also possible that MW firms that choose to improve their internal controls can introduce a possible self-selection bias into our tests of H2. To control for this potential bias, we use the Heckman two-stage procedure and estimate a probit regression of *FIXED* on the same determinants of MWs as above, reasoning that the characteristics that lead to the incidence of MWs are also likely to cause difficulties for these firms to remediate MWs. We also include governance characteristics because firms with better governance are more likely to remediate MWs (Goh 2009).[23] We hand-collect data on the board and audit committee characteristics from the proxy statements of the 516 MW firms used earlier to test H2, and are left with 437 firms with available data.

Table 8 reports the results after including the inverse Mills ratio, *LAMBDA*, in our previous regressions used to test H2. For brevity, we omit the results for the control variables. The results in Table 8 are similar to those when we do not control for self-selection. Specifically, the coefficients on $D\Delta NI_{t-1} * \Delta NI_{t-1} * FIXED$, $DCFO * CFO * FIXED$, and $DR * R * FIXED$ remain significant and have the same signs as those in Table 6. We continue to find similar significant results (untabulated) when we include the determinants of *FIXED*, instead of *LAMBDA* as additional control variables in the regressions. In sum, we conclude that the results for H2 are unlikely to be driven by self-selection in the improvement in internal control quality.

Alternative Measures of Conditional Conservatism

Khan and Watts (2009) suggest a firm-specific measure of conservatism (*C_SCORE*) that is based on Basu's (1997) measure of asymmetric timeliness of earnings. We follow their approach and derive the *C_SCORE* measure, with higher *C_SCORE* indicating greater conservatism. The regression model and results are presented in Table 9.[24] In summary, the results using the *C_SCORE* measure confirm our earlier findings. Specifically, the coefficient on *MW* in Column 1

---

[22] To save space, we do not tabulate the results. However, these results, as well as all other results that we do not tabulate, are available from the authors upon request.

[23] These governance characteristics include board size, board independence, board meeting frequency, audit committee size, audit committee accounting financial expertise, audit committee meeting frequency, the duality of the CEO and Chairman positions, board commitment (proxied by the number of outside directorships in public companies held by the board directors), and board diligence (proxied by the proportion of directors who fail to attend more than 75 percent of the board meetings). The first-stage regression results show that firm profitability, the occurrence of a restatement, board independence, and audit committee meeting frequency are significantly associated with the likelihood of the remediation of MWs. Details on the definitions of the governance characteristics and the first-stage regression results are available from the authors upon request.

[24] Khan and Watts (2009) show that firms with high uncertainty and long investment cycles have future gains that are less verifiable and are more susceptible to gaming, generating a higher contracting demand for conservatism. Hence, we control for information uncertainty using *STD* (defined as the standard deviation of monthly stock returns) and control for the length of the investment cycle using *CYCLE* (defined as the depreciation expenses deflated by lagged asset). We control for *AGE* (firm age in years) because younger firms tend to have larger information asymmetry that is associated with more conservatism. Finally, we include *SPREAD* (the bid-ask spread scaled by the midpoint of the spread) to proxy for information asymmetry that is positively related to conservatism (LaFond and Watts 2008).

TABLE 8

Regression Results of H2 after Controlling for Self-Selection of Remediation of MWs Using the Heckman Two-Stage Procedure

	Pred. Sign	Persistence of Earnings Changes $A = D\Delta NI_{t-1}$ $B = \Delta NI_{t-1}$ (1) Coeff.	p-value	Accrual-Based Loss Recognition $A = DCFO$ $B = CFO$ (2) Coeff.	p-value	Timeliness of Earnings to News $A = DR$ $B = R$ (3) Coeff.	p-value
Intercept	?	0.004	0.88	−0.010	0.65	0.014	0.89
A	?	0.000	0.99	−0.102***	<0.01	0.100	0.41
B	?	0.522***	<0.01	−0.496***	<0.01	0.132	0.45
*A * B*	−, +, +	0.230	0.37	0.457*	0.09	0.455	0.11
FIXED	?	0.009	0.73	0.005	0.78	0.049	0.63
*A * FIXED*	?	−0.001	0.98	0.088***	<0.01	0.166	0.24
*B * FIXED*	?	0.348	0.13	0.055	0.69	−0.031	0.87
*A * B * FIXED*	?	−0.905**	0.01	0.664***	<0.01	0.893**	0.01
LAMBDA	?	−0.021	0.35	−0.042***	<0.01	−0.144**	0.05
Control Variables		Not reported		Not reported		Not reported	
Industry Effects		Not reported		Not reported		Not reported	
Adjusted R^2		24.6%		49.4%		23.1%	
Number of observations		437		437		437	

*, **, *** Denote significance at the 10 percent, 5 percent, and 1 percent levels, respectively, based on two-tailed tests.

This table reports the results of the regressions that control for self-selection in the remediation of MWs. The results are from the second-stage regression of the Heckman (1979) two-stage procedure, after controlling for *LAMBDA*. *LAMBDA* is the inverse Mill ratio calculated from the first-stage regression in which we regress *FIXED* on the determinants of MW remediation. The determinants of MW remediation include the determinants of MWs and the governance characteristics detailed in footnote 23. For brevity, we do not report the results for the control variables. Columns 1, 2, and 3 present the results using the persistence of earnings changes measure, accrual-based loss recognition measure, and the timeliness of earnings to news measure, respectively. In Column 1, the dependent variable is ΔNI_t, which is the change in net income before extraordinary items for firm *i* in fiscal year *t* deflated by beginning-of-year total assets. In Column 2, the dependent variable is *ACCRUAL*, which is net income before extraordinary items minus operating cash flows for firm *i* in fiscal year *t* deflated by beginning-of-year total assets. In Column 3, the dependent variable is *NI*, which is the net income before extraordinary items for firm *i* in fiscal year *t* deflated by the beginning-of-year market value of equity. All the other variables are as previously defined.

is −0.007 (p = 0.02), consistent with MW firms exhibiting lower levels of conservatism than the control firms. The coefficient on *FIXED* in Column 2 is 0.288 (p = 0.03), consistent with MW firms that remediate their weaknesses exhibiting greater conservatism than MW firms that continue to have weaknesses.

To further corroborate our findings, we use a second accrual-based measure of conservatism (*CON-ACC*) that is suggested by Givoly and Hayn (2000). This measure is computed as income before extraordinary items less cash flows from operations plus depreciation expense deflated by average total assets, averaged over a three-year period before year *t*.[25] The lower the *CON-ACC*,

[25] The intuition behind this measure is that conservative accounting results in persistently negative accruals. Averaging the accruals over a number of periods ensures that the effects of any temporary large accruals are mitigated, as accruals tend to reverse within a one- to three-year period (Richardson et al. 2005). For the H1 test, we measure *CON-ACC* over the

TABLE 9

Regression Results Using the Firm-Specific Conservatism Measure in Khan and Watts (2009)

H1:

$$C\_SCORE = \lambda_0 + \lambda_1 MW + \lambda_2 SIZE + \lambda_3 LEV + \lambda_4 MB + \lambda_5 STD + \lambda_6 CYCLE + \lambda_7 AGE + \lambda_8 SPREAD + \varepsilon.$$

H2:

$$C\_SCORE = \lambda_0 + \lambda_1 FIXED + \lambda_2 SIZE + \lambda_3 LEV + \lambda_4 MB + \lambda_5 STD + \lambda_6 CYCLE + \lambda_7 AGE + \lambda_8 SPREAD + \varepsilon.$$

	Pred. Sign	Test of H1 (1) Coeff.	Test of H1 (1) p-value	Test of H2 (2) Coeff.	Test of H2 (2) p-value
Intercept	?	0.484***	<0.01	1.408*	0.07
MW	?	−0.007**	0.02		
FIXED	?			0.288**	0.03
SIZE	−	−0.067***	<0.01	−0.184	0.14
LEV	+	0.233***	<0.01	1.847***	<0.01
MB	?	−0.001**	0.02	−0.025**	0.03
STD	+	0.063***	<0.01	1.373	0.26
CYCLE	−	−0.340***	<0.01	−6.569***	<0.01
AGE	−	0.000	0.64	−0.008	0.19
SPREAD	+	−0.011**	0.02	2.086***	0.01
Industry Effects		Not reported		Not reported	
Adjusted R2		73.8%		7.0%	
Number of observations		5836		500	

*, **, *** Denote significance at the 10 percent, 5 percent, and 1 percent levels, respectively, based on two-tailed tests.
This table reports the results of the regressions that examine H1 and H2 using the firm-specific conservatism measure *C_SCORE* in Khan and Watts (2009). The dependent variable is *C_SCORE*, with higher *C_SCORE* indicating greater conservatism. The other variables are defined as follows or as previously defined.

Variable Definitions:

STD = standard deviation of monthly stock returns;
CYCLE = depreciation expenses deflated by the beginning-of-year total assets, which is a decreasing measure of the length of the investment cycle;
AGE = firm age, the number of years since the firm appears in CRSP database; and
SPREAD = bid-ask spread scaled by the midpoint of the spread, obtained from CRSP.

three-year period before each 2000 and 2001 firm-year observation. For the H2 test, we measure *CON-ACC* over the three-year period right after the year of the second SOX 404 report. In these tests, we control for firm size (*SIZE*), leverage (*LEV*), and litigation risks (*LIT*) as before. Consistent with Ahmed et al. (2002), we control for growth in sales (*SALESGROWTH*) because it will affect accruals such as inventory and receivables, which in turn affect *CON-ACC*. We control for research and development plus advertising expenditures (*RDADV*) because they are likely to capture economic rents generated by assets-in-place, growth opportunities, and GAAP mandated conservatism (Ahmed 1994). Finally, we control for profitability (*CFO*) as proxied by the operating cash flows deflated by beginning-of-year total assets, because Ahmed et al. (2002) argue that profitable firms tend to use more conservative accounting.

the more conservative the firm's accounting. Table 10 presents the results using *CON-ACC*, which is consistent with all our earlier findings. Specifically, the coefficient on *MW* in Column 1 is 0.010 (p = 0.05), implying that MW firms have larger operating accruals (i.e., lower conservatism) than the control firms. Further, the coefficient on *FIXED* in Column 2 is −0.005 (p = 0.08), which indicates that MW firms that remediate their weaknesses exhibit greater conservatism than MW firms that fail to do so. In sum, our earlier findings on H1 and H2 are robust to alternative conservatism measures suggested by Khan and Watts (2009) and Givoly and Hayn (2000).

Expanded Sample for H2 Tests

In our H2 tests in Table 6, we remove 473 firms without SOX 404 opinions; that is, firms that are either non-accelerated filers (289 firms) or have terminated the registration of their securities

TABLE 10

Regression Results Using the Accrual-Based Conservatism Measure in Givoly and Hayn (2000)

H1:

$$CON\text{-}ACC = \theta_0 + \theta_1 MW + \theta_2 SIZE + \theta_3 SALESGROWTH + \theta_4 RDADV + \theta_5 CFO + \theta_6 LEV + \theta_7 LIT + \varepsilon.$$

H2:

$$CON\text{-}ACC = \theta_0 + \theta_1 FIXED + \theta_2 SIZE + \theta_3 SALESGROWTH + \theta_4 RDADV + \theta_5 CFO + \theta_6 LEV + \theta_7 LIT + \varepsilon.$$

		Test of H1 (1)		Test of H2 (2)	
	Pred. Sign	Coeff.	p-value	Coeff.	p-value
Intercept	?	−0.079***	<0.01	−0.021***	<0.01
MW	?	0.010**	0.05		
FIXED	?			−0.005*	0.08
SIZE	+	0.007***	<0.01	0.002**	0.03
SALESGROWTH	+	−0.003	0.38	0.005*	0.10
RDADV	+	−0.012***	<0.01	0.000	0.63
CFO	−	0.068***	<0.01	0.052***	<0.01
LEV	−	−0.009	0.45	−0.010	0.18
LIT	−	−0.029***	<0.01	−0.008**	0.04
Industry Effects		Not reported		Not reported	
Adjusted R2		3.5%		10.4%	
Number of observations		7,634		407	

*, **, *** Denote significance at the 10 percent, 5 percent, and 1 percent levels, respectively, based on two-tailed tests.

This table reports the results of the regressions that examine H1 and H2 using the accrual-based conservatism measure in Givoly and Hayn (2000). The dependent variable is *CON-ACC*, which is income before extraordinary items less cash flows from operations plus depreciation expense deflated by average total assets, and averaged over a three-year period. The lower the *CON-ACC*, the greater the conservatism. The other variables are defined as follows or as previously defined.

Variable Definitions:

SALESGROWTH = percentage of annual growth in total sales (Data12); and

RDADV = research and development expenditures plus advertising expense divided by sales.

(184 firms). This procedure results in a loss of roughly half of our *MW*s observations. To recover some of these lost observations, we relax our requirements to determine the remediation of *MW*s by using the SOX 302 opinions for non-accelerated filers. For these non-accelerated filers, we rely on the SOX 302 opinions at the time of the second SOX 404 report that would have been issued if the firm had been an accelerated filer. This procedure yields larger sample sizes of 814, 800, and 651 firms with available data from Compustat and/or CRSP for the persistence of earnings changes measure, accruals-based loss recognition measure, and timeliness of earnings to news measure, respectively.[26] We then replicate the H2 tests in Table 6 using these expanded samples, and find that the inferences of our earlier results remain unchanged. Specifically, untabulated results show that the coefficients on $D\Delta NI_{t-1} * \Delta NI_{t-1} * FIXED$, $DCFO * CFO * FIXED$, and $DR * R * FIXED$ are -1.393 ($p < 0.01$), 0.557 ($p = 0.06$), and 0.607 ($p = 0.01$), respectively.

Contextual Analysis of the Nature of MWs

Finally, to provide a stronger link between internal controls and conservatism, we further explore the nature of the MWs. Using the arguments found in Ball (2001), we expect firms with MWs pertaining to investment and/or asset management related decisions to have a greater incentive to avoid timely loss recognition.[27] In addition, following the arguments in Doyle et al. (2007a), we expect company-level MWs to be more severe and pervasive than other types of MWs, and consequently to result in even lower conservatism.[28] Hence, we code firms with at least one of the above types of MWs as "severe MW firms" and test whether these firms exhibit lower conservatism than the other MW firms. To conduct the analyses, we replace MW with *SEVERE* in the regression models (2), (4), and (6), where *SEVERE* is an indicator variable that equals 1 for severe MW firms, and 0 otherwise. Using the MW firms observations from fiscal years 2001 and 2002, the regression results (untabulated) reveal that the coefficients on $D\Delta NI_{t-1} * \Delta NI_{t-1} * SEVERE$, $DCFO * CFO * SEVERE$, and $DR * R * SEVERE$ are 0.677 ($p < 0.01$), -0.069 ($p = 0.57$), and -0.127 ($p = 0.07$), respectively. Hence, the contextual analysis of the nature of MWs shows some evidence that severe MW firms exhibit lower conservatism than the other MW firms. This evidence provides further empirical support of a positive relation between internal control quality and conservatism.

---

[26] The significantly smaller sample size for the timeliness of earnings to news measure is due to the additional requirements for returns data from CRSP. Unlike before, we do not require that all three tests using different measures of conservatism have the same sample size. In other words, we do not impose the same data limitations to all measures of conservatism. This allows us to obtain the maximum sample size for each measure of conservatism to conduct the additional analyses on H2.

[27] Ball (2001) states that both debt contracting and corporate governance give rise to the demand for accounting conservatism, because conservatism can incorporate into financial statements adverse information about future cash flows of investment that is observable only to managers. MWs in the investment and/or asset management-related decisions can impede this governance role of conservatism. For instance, control weaknesses over the computation of the future cash flows of acquired investments and the lack of staff expertise in estimating future cash flows for asset impairment tests make it harder for the accounting system to incorporate into financial statements adverse information about future cash flows, which will lead to lower level of conservatism. The examples of MWs found in the Appendix are illustrative of these types of MWs.

[28] Following Doyle et al. (2007a), company-level MWs include override by senior management and ineffective control environment. These MWs are more severe and pervasive than account-specific MWs, which include inadequate internal controls for accounting for loss contingencies, deficiencies in the documentation of a receivables securitization program, and inadequate internal controls over the application of new accounting principles or the application of existing accounting principles to new transactions. Consistent with Doyle et al. (2007a), we also consider firms that have at least three account-specific MWs to have company-level MWs. We would like to thank Jeffrey Doyle, Weili Ge, and Sarah McVay for sharing the data on the classification of MWs.

VI. CONCLUSION

Our results show that firms that disclose material weaknesses (MWs) exhibit lower conditional conservatism than firms without such weaknesses, and MW firms that remediate their weaknesses exhibit greater conservatism than MW firms that fail to do so. Collectively, our results are consistent with strong internal controls facilitating conservatism. These findings extend the literature on the reporting effects of strong versus weak internal controls. In addition, our study contributes to the broader literature that tests the extent to which conservatism is shaped by a firm's contracting and governance environment.

One limitation of our study is that for our tests of H2, we remove nearly half of the 1,098 MW firms that do not have SOX 404 opinions due to their non-accelerated filer status and due to termination of securities registration. It is unclear how the addition of these firms would affect the results of H2. Another limitation is that our findings cannot directly shed light on the current debate regarding the costs and benefits of SOX. Our study informs only of the value of internal control quality, but not of the effectiveness of the internal control reporting requirements under SOX.

APPENDIX
EXAMPLES OF MATERIAL WEAKNESSES THAT COULD IMPEDE THE TIMELY RECOGNITION OF LOSSES

NEOMAGIC CORP (28 April, 2005)

NeoMagic Corporation has a material weakness as of January 31, 2005 related to its process of evaluating long-lived assets for impairment, including insufficient controls over the review and documentation of key assumptions used in preparing forecasted cash flows used to support NeoMagic Corporation's impairment analyses. This material weakness resulted in an audit adjustment that was recorded by NeoMagic Corporation to recognize the impairment of intangible assets.

OPENTV CORP (16 March, 2005)

We did not have sufficient internal personnel and technical expertise to properly apply accounting principles to certain non-routine matters; in particular, we applied certain provisions of Statement of Financial Accounting Standards No. 142, *Goodwill and Other Intangibles* (SFAS No. 142), in an incorrect manner when conducting our annual analysis for potential impairment of goodwill, which led to an error that was identified and subsequently corrected.

CLEARONE COMMUNICATIONS INC (18 August, 2005)

We have a material weakness related to the tracking and valuation of inventory, including controls to identify and properly account for obsolete inventory. Our accounting policies and practices over tracking and valuation of inventory, including controls to identify and properly account for slow-moving, obsolete inventory were inconsistent with GAAP. This material weakness resulted in errors in recording inventories at the lower of cost or market, and errors for inventory shrinkage.

We have a material weakness in accounting for non-routine transactions, which include business combinations, discontinued operations, sale of a business unit (other than discontinued operations), and evaluation and recognition of impairment charges. This material weakness resulted in improper purchase price allocations in business combinations, improper amortization and depreciation of long-lived assets, improper identification and recording of activities related to discontinued operations, improper recording and reporting the sale of business units, improper evaluation of triggering events associated with impairment of long-lived assets (including annual impairment tests for goodwill); and improper calculating and recording of impairment charges.

ASCONI CORP (27 May 2004)

We lacked controls to ensure that agreements, contracts, and other documents relating to investments in subsidiaries or investees were provided to and reviewed by accounting and financial reporting personnel on a timely basis to ensure that the financial reporting and disclosure implications of such transactions could be considered and reflected in the financial statements in the proper periods.

FINANCIAL INDUSTRIES CORP (10 June, 2003)

Management identified a significant number of internal control weaknesses in several key areas that had resulted in material misstatement of financial results. These include weaknesses in the following areas, among others: deferred policy acquisition costs; present value of future profits of acquired businesses; investment accounting; consolidation process; purchase accounting; and intercompany accounting.

REFERENCES

Ahmed, A. S. 1994. Accounting earnings and future economic rents: An empirical analysis. *Journal of Accounting and Economics* 17 (3): 377–400.

———, B. K. Billings, R. M. Morton, and M. Stanford-Harris. 2002. The role of accounting conservatism in mitigating bondholder-shareholder conflicts over dividend policy and in reducing debt costs. *The Accounting Review* 77 (4): 867–890.

———, and S. Duellman. 2007. Accounting conservatism and board of director characteristics: An empirical analysis. *Journal of Accounting and Economics* 43 (2–3): 411–437.

Aiken, L., and S. West. 1991. *Multiple Regression: Testing and Interpreting Interactions*. Thousand Oaks, CA: SAGE Publications Inc.

Armstrong, C. S., W. R. Guay, and J. P. Weber. 2010. The role of information and financial reporting in corporate governance and debt contracting. Working paper, University of Pennsylvania.

Ashbaugh-Skaife, H., D. Collins, and W. Kinney. 2007. The discovery and reporting of internal control deficiencies prior to SOX-mandated audits. *Journal of Accounting and Economics* 44 (1–2): 166–192.

———, ———, ———, and R. LaFond. 2008. The effect of SOX internal control deficiencies and their remediation on accruals quality. *The Accounting Review* 83 (1): 217–250.

———, ———, ———, and ———. 2009. The effect of SOX internal control deficiencies on firm risk and cost of equity. *Journal of Accounting Research* 47 (1): 1–43.

Ball, R., 2001. Infrastructure requirements for an economically efficient system of public financial reporting and disclosure. *Brookings-Wharton Papers on Financial Services*: 127–169.

———, S. P. Kothari, and A. Robin. 2000. The effect of international institutional factors on properties of accounting earnings. *Journal of Accounting and Economics* 29 (1): 1–51.

———, A. Robin, and J. Wu. 2003. Incentives versus standards: Properties of accounting income in four East Asian countries. *Journal of Accounting and Economics* 36 (1–3): 235–270.

———, and L. Shivakumar. 2005. Earnings quality in UK private firms: Comparative loss recognition timeliness. *Journal of Accounting and Economics* 39 (1): 83–128.

———, and ———. 2006. The role of accruals in asymmetrically timely gain and loss recognition. *Journal of Accounting Research* 44 (2): 207–242.

Barth, M. 2008. Global financial reporting: Implications for U.S. academics. *The Accounting Review* 83 (5): 1159–1179.

Basu, S. 1997. The conservatism principle and the asymmetric timeliness of earnings. *Journal of Accounting and Economics* 24 (1): 3–37.

———, L. Hwang, and C. Jan. 2001. Differences in conservatism between Big Eight and non-Big Eight auditors. Working paper, Emory University, Chinese University of Hong Kong, and California State University, Hayward.

Beaver, W. H., and S. G. Ryan. 2005. Conditional and unconditional conservatism: Concepts and modeling.

Review of Accounting Studies 10 (2–3): 269–309.

Beneish, M. D., M. Billings, and L. Hodder. 2008. Internal control weaknesses and information uncertainty. *The Accounting Review* 83 (3): 665–703.

Bushman, R., and A. Smith. 2001. Financial accounting information and corporate governance. *Journal of Accounting and Economics* 32 (1–3): 237–333.

———, and J. Piotroski. 2006. Financial reporting incentives for conservative accounting: The influence of legal and political institutions. *Journal of Accounting and Economics* 42 (1–2): 107–148.

Collins, D., E. Maydew, and I. Weiss. 1997. Changes in the value-relevance of earnings and book values over the past forty years. *Journal of Accounting and Economics* 24 (1): 39–67.

Committee of Sponsoring Organizations of the Treadway Commission (COSO). 1992. *Internal Control—Integrated Framework.* The framework. New York, NY: COSO.

Dietrich, J., K. Muller, and E. Riedl. 2007. Asymmetric timeliness tests of accounting conservatism. *Review of Accounting Studies* 12 (1): 95–124.

Doyle, J., W. Ge, and S. McVay. 2007a. Accruals quality and internal control over financial reporting. *The Accounting Review* 82 (5): 1141–1170.

———, ———, and ———. 2007b. Determinants of weaknesses in internal control over financial reporting. *Journal of Accounting and Economics* 44 (1–2): 193–223.

Ely, K., and G. Waymire. 1999. Accounting standard-setting organizations and earnings relevance: Longitudinal evidence from NYSE common stocks, 1927–1973. *Journal of Accounting Research* 37 (2): 293–317.

Fama, E. F., and M. C. Jensen. 1983. Separation of ownership and control. *The Journal of Law & Economics* 26 (2): 301–325.

Feng, M., C. Li, and S. McVay. 2009. Internal control and management guidance. *Journal of Accounting and Economics* 48 (2–3): 190–209.

Financial Accounting Standards Board (FASB). 1980. *Qualitative Characteristics of Accounting Information.* Statement of Financial Accounting Concepts No. 2. Norwalk, CT: FASB.

———. 2000. *Using Cash Flow Information and Present Value in Accounting Measurements.* Statement of Financial Accounting Concepts No. 7. Norwalk, CT: FASB.

———. 2008. *Exposure Draft, Conceptual Framework for Financial Reporting: The Objective of Financial Reporting and Qualitative Characteristics and Constraints of Decision-Useful Financial Reporting Information.* May 28. Norwalk, CT: FASB.

Francis, J. R., D. R. Philbrick, and K. Schipper. 1994. Shareholder litigation and corporate disclosures. *Journal of Accounting Research* 32 (2): 137–164.

———, and K. Schipper. 1999. Have financial statements lost their relevance? *Journal of Accounting Research* 37 (2): 319–352.

Frankel, R. M., M. F. Johnson, and K. K. Nelson. 2002. The relation between auditors' fees for non-audit services and earnings management. *The Accounting Review* 77 (Supplement): 71–106.

Givoly, D., and C. Hayn. 2000. The changing time-series properties of earnings, cash flows and accruals: Has financial reporting become more conservative? *Journal of Accounting and Economics* 29 (3): 287–320.

———, ———, and A. Natarajan. 2007. Measuring reporting conservatism. *The Accounting Review* 82 (1): 65–106.

———, ———, and S. P. Katz. 2010. Does public ownership of equity improve earnings quality? *The Accounting Review* 85 (1): 195–225.

Goh, B. W. 2009. Audit committees, boards of directors, and remediation of material weaknesses in internal control. *Contemporary Accounting Research* 26 (2): 549–579.

Guay, W., and R. Verrecchia. 2006. Discussion of: The role of accruals in asymmetrically timely gain and loss recognition. *Journal of Accounting Research* 44 (2): 243–255.

Heckman, J. J. 1979. Sample selection bias as a specification error. *Econometrica* 47 (1): 153–162.

Holthausen, R. W., and R. L. Watts. 2001. The relevance of value-relevance literature for financial accounting standard-setting. *Journal of Accounting and Economics* 31 (1–3): 3–75.

International Accounting Standards Board (IASB). 2001. *Framework for the Preparation and Presentation of Financial Statements.* London, UK: IASB.

Jensen, M. C. 1993. The modern industrial revolution, exit and the failure of internal control systems. *The Journal of Finance* 48 (3): 831–880.

Khan, M., and R. L. Watts. 2009. Estimation and empirical properties of a firm-year measure of conservatism. *Journal of Accounting and Economics* 48 (2–3): 132–150.

LaFond, R., and S. Roychowdhury. 2008. Managerial ownership and accounting conservatism. *Journal of Accounting Research* 46 (1): 101–135.

———, and R. L. Watts. 2008. The information role of conservative financial statements. *The Accounting Review* 83 (2): 447–478.

LaLonde, R. 1986. Evaluating the econometric evaluations of training programs with experimental data. *The American Economic Review* 76 (4): 604–620.

Leuz, C., and R. Verrecchia. 2000. The economic consequences of increased disclosure. *Journal of Accounting Research* 38 (Supplement): 91–124.

Levitt, A. 1998. The numbers game. Remarks delivered at the NYU Center for Law and Business, New York, NY, September 28.

Neter, J., W. Wasserman, and M. H. Kutner. 1989. *Applied Linear Regression Models*. 2nd edition. Homewood, IL: Irwin.

Nikolaev, V. 2010. Debt covenants and accounting conservatism. *Journal of Accounting Research* 48 (1): 51–89.

Ogneva, M., K. R. Subramanyam, and K. Raghunandan. 2007. Internal control weaknesses and cost of equity: Evidence from SOX Section 404 disclosures. *The Accounting Review* 82 (5): 1255–1297.

Penman, S., and X. Zhang. 2002. Accounting conservatism, the quality of earnings and stock returns. *The Accounting Review* 77 (2): 237–264.

Public Company Accounting Oversight Board (PCAOB). 2004. *An Audit of Internal Control Over Financial Reporting Performed in Conjunction with an Audit of Financial Statements*. Auditing Standards No. 2. Washington, D.C.: PCAOB.

Richardson, S. A., R. G. Sloan, M. T. Soliman, and I. A. Tuna. 2005. Accrual reliability, earnings persistence and stock prices. *Journal of Accounting and Economics* 39 (3): 437–485.

Roychowdhury, S., and R. Watts. 2007. Asymmetric timeliness of earnings, market-to-book and conservatism in financial reporting. *Journal of Accounting and Economics* 44 (1–2): 2–31.

Ryan, S. G. 2006. Identifying conditional conservatism. *European Accounting Review* 15 (4): 511–525.

Securities and Exchange Commission (SEC). 2002. *Final Rule: Certification of Disclosure in Companies' Quarterly and Annual Reports*. Release Nos. 33-8124; 34-46427. Washington, D.C.: SEC

———. 2003. *Final Rule: Management's Reports on Internal Control over Financial Reporting and Certification of Disclosure in Exchange Act Periodic Reports*. Release Nos. 33-8238; 34-47986. Washington D.C.: SEC August 14.

———. 2004. *Management's Reports on Internal Control over Financial Reporting and Certification of Disclosure in Exchange Act Periodic Reports-Frequently Asked Questions*. Washington D.C.: SEC.

Watts, R. L. 2003. Conservatism in accounting part 1: Explanations and implications. *Accounting Horizons* 17 (3): 207–221.

———, and J. L. Zimmerman. 1986. *Positive Accounting Theory*. Englewood Cliffs, NJ: Prentice Hall Inc.

Zhang, J. 2008. The contracting benefits of accounting conservatism to lenders and borrowers. *Journal of Accounting and Economics* 45 (1): 27–54.

THE ACCOUNTING REVIEW
Vol. 86, No. 3
2011
pp. 1007–1043

American Accounting Association
DOI: 10.2308/accr.00000042

Relative Performance Evaluation and Related Peer Groups in Executive Compensation Contracts

Guojin Gong
The Pennsylvania State University

Laura Yue Li
University of Illinois at Urbana–Champaign

Jae Yong Shin
Seoul National University

ABSTRACT: This study examines the explicit use of relative performance evaluation (RPE) in executive compensation contracts and the selection of RPE peers. Using S&P 1500 firms' first proxy disclosures under the SEC's 2006 executive compensation disclosure rules, we find that about 25 percent of our sample firms explicitly use RPE in setting executive compensation. We demonstrate that a lack of knowledge of both actual peer-group composition and the link between RPE-based performance targets and future peer performance significantly hinder the traditional implicit test from detecting RPE use. We also find that firms consider both costs and benefits of RPE as an incentive mechanism when deciding to use RPE. Finally, both efficient contracting and rent extraction considerations influence RPE peer selection, with the relative importance of these competing considerations depending on RPE firms' performance.

Keywords: *executive compensation; relative performance evaluation; peer group; SEC regulation.*

Data Availability: *Data are available from public sources indicated in the text.*

We thank Ana Albuquerque, Mary Ellen Carter, Ingolf Dittmann, Ron Dye, Zhaoyang Gu, Ranjani Krishnan (editor), Ronald Masulis, two anonymous referees, and workshop participants at the 6th Accounting Research Workshop at the Universities of Fribourg and Bern, the 2009 Chinese Accounting Professors' Association of North America Meeting, Northwestern University, The Pennsylvania State University, University of Illinois at Urbana–Champaign, University of Kentucky, University of Wisconsin–Madison, and Vanderbilt University for their helpful comments. We also thank Jean Liang, Maria Revelo, Tony Shin, Jeff Wang, Steve Wu, and Victor Yang for their valuable research assistance. Professor Shin acknowledges financial support from the Institute of Management Research at Seoul National University.

Editor's note: Accepted by Ranjani Krishnan.

Submitted: March 2009
Accepted: October 2010
Published Online: April 2011

I. INTRODUCTION

Relative performance evaluation (RPE) entails the use of peer performance in evaluating an agent's performance. Based upon the premise that peer performance captures common exogenous shocks, the major theoretical benefit of RPE is to insulate an agent from common risk, while providing the principal with a more informative measure to assess the agent's actions. Consequently, RPE is expected to lead to risk-sharing benefits in incentive contracting (e.g., Lazear and Rosen 1981; Holmstrom 1982; Nalebuff and Stiglitz 1983).

Despite RPE's theoretical appeal, prior empirical research offers mixed evidence on the use of RPE in executive compensation contracts based on an *implicit* approach.[1] The implicit approach infers RPE use by regressing executive pay on industry performance across a population of firms, and thus relies on simplified assumptions concerning RPE contract details (such as RPE peer-group composition, performance metrics used in RPE, and components of pay covered by RPE). These assumptions unavoidably introduce measurement errors into implicit tests (Murphy 1999; Bannister and Newman 2003). More importantly, studies that rely on the implicit approach are unable to examine the execution of RPE contracts (such as peer selection), which is essential to evaluating the costs and benefits of RPE as an incentive mechanism (Matsumura and Shin 2006; Bol et al. 2010).

The SEC's 2006 executive compensation disclosure rules require firms to provide details on how performance targets (including *relative* performance targets) are used in setting executive pay. We exploit this mandatory disclosure regime to examine the explicit use of RPE and related peer groups. We first examine the extent to which simplified assumptions concerning RPE contract details affect inferences drawn from the traditional implicit approach. We then examine firms' decisions to explicitly use RPE in setting executive pay. Finally, we focus on RPE peer groups, an essential element of RPE contracts, to investigate factors that might influence the selection of peers for RPE purposes.

Using S&P 1500 firms' first proxy disclosures under the SEC's new disclosure rules, we find that about 25 percent of S&P 1500 firms explicitly use RPE in setting executive compensation. When using the implicit approach (such as in Albuquerque [2009]), where RPE peers are matched on both industry and size, we do not find evidence of RPE use in S&P 1500 firms in our 2006 sample.[2] We further show that this implicit approach is unable to detect RPE use even among firms that claim to use RPE in setting executive pay. However, after incorporating disclosed peer-group composition, we find a significantly negative association between CEO pay and stock performance for disclosed peers, supporting the use of RPE. Moreover, we find that failure to account for the link between RPE-based performance targets and future peer performance hinders the detection of RPE use in implicit tests. These results illustrate that simplified assumptions about RPE contract details cloud inferences drawn from implicit tests.

Turning to firms' decisions to explicitly use RPE, we find that firms exposed to higher common risk, operating in less-concentrated industries, having fewer growth opportunities, and hiring less-wealthy CEOs are more likely to use RPE. These findings are in line with economic theories predicting that a firm's use of RPE increases with the firm's exposure to common risk and decreases with the CEO's self-hedging ability (e.g., Lazear and Rosen 1981; Holmstrom 1982; Garvey and Milbourn 2003). Moreover, firms that are larger, have more independent and larger boards, and hire compensation consultants are more likely to use RPE. These results reveal the

[1] Examples include Antle and Smith (1986), Barro and Barro (1990), Gibbons and Murphy (1990), Janakiraman et al. (1992), Joh (1999), Aggarwal and Samwick (1999), Garvey and Milbourn (2003), Rajgopal et al. (2006), and Albuquerque (2007, 2009).

[2] In untabulated results, using the implicit approach as per Albuquerque (2009), we replicate Albuquerque's main finding and obtain a significantly negative coefficient on industry-size-matched peer returns over the 1992–2005 period.

importance of board structure and compensation consultants in facilitating the use of RPE. The overall evidence supports the view that firms consider both costs and benefits of RPE as an incentive mechanism when deciding to use RPE.

Next, among firms using RPE in the pay-setting process, we examine whether firms choose RPE peers in a manner consistent with efficient contracting or rent extraction. Under the efficient-contracting perspective, the selection of RPE peers should consider potential peers' ability to capture common risk (e.g., Holmstrom 1982). Tournament theory further suggests selecting RPE peers such that RPE firms' and peer firms' executives have equal chances to "win," given similar effort levels (e.g., Lazear and Rosen 1981). Under the rent-extraction perspective, executives have incentives to choose RPE peers that they expect to underperform (e.g., Gibbons and Murphy 1990; Murphy 2001), as choosing peers with poor expected performance can inflate RPE firms' peer-adjusted performance and, hence, increase executive compensation.

The evidence supports both efficient contracting and rent-extraction behavior in the peer-selection process. Consistent with efficient contracting, RPE firms tend to choose peers with higher common-risk-reduction ability and with greater similarity. These findings are generally more pronounced for RPE firms that outperform their industry than for underperforming RPE firms. At the same time, we find that RPE firms are more likely to select peers that are expected to perform poorly (based on analysts' target price forecasts). Interestingly, this self-serving bias in peer selection is more significant among underperforming RPE firms. Taken together, our findings suggest that RPE firms' superior performance leads to efficient-contracting considerations, whereas RPE firms' underperformance leads to rent-seeking behavior in the peer-selection process. Executives in underperforming RPE firms likely have stronger incentives to behave opportunistically in the peer-selection process because they are more concerned about potential reputation loss and pay reduction as a result of poor performance. When we examine intertemporal changes in peer-group composition, we find some evidence that added peers are more similar in size to RPE firms, but exhibit poorer performance than dropped peers.

Our study contributes to the executive compensation literature in three ways. First, our evidence facilitates a better understanding of the "RPE puzzle" that has long plagued researchers (Murphy 1999; Prendergast 1999). Our findings suggest that the RPE puzzle of mixed empirical support for RPE use arises at least in part from simplified assumptions underlying the traditional implicit approach. We demonstrate that incorporating RPE contract details (such as peer-group composition and the link between RPE-based performance targets and future peer performance) can significantly alter inferences drawn from implicit tests. Our evidence thus cautions against relying on the traditional implicit approach to draw definitive conclusions regarding RPE use.

Second, we provide the first large-sample U.S. evidence on the explicit use of RPE. We find that about 25 percent of our sample firms use RPE in setting executive compensation. Our findings on firms' decisions to use RPE confirm related economic theories, while also revealing the importance of board structure and compensation consultants in facilitating RPE use. This stands in contrast to Carter et al. (2009), who do not find evidence in support of related economic theories when studying explicit RPE use in performance-vested equity plans for a sample of U.K. firms.

Third, we provide the first large-sample evidence on the selection of RPE peers. We document that both efficient contracting (captured by common risk and economic similarity) and rent extraction (measured by expected peer performance) explain the likelihood of being selected as a RPE peer. These findings shed light on the execution of RPE contracts, which is essential to understanding the overall costs and benefits of RPE as an incentive device.

Our evidence also adds to the growing literature on the use of peer groups in executive compensation contracts. Existing research predominantly focuses on peer groups used for compensation-benchmarking purposes (Bizjak et al. 2008, 2010; Cadman and Carter 2010; Albuquerque et al. 2009; Faulkender and Yang 2010). The purpose of compensation benchmarking

is to gauge the appropriate *level* of compensation, while the purpose of RPE is mainly to employ external performance targets (i.e., peer performance) to filter out common risk. Selecting better (worse) performing firms as benchmarking (RPE) peers is expected to increase executive compensation. In additional analysis, we find that, for firms disclosing both benchmarking peers and RPE peers, peers used *solely* for RPE (benchmarking) purposes exhibit higher (lower) performance comovement with the firms of interest and have worse (better) stock performance than unselected peers. These contrasting results indicate different considerations in selecting RPE peers versus compensation-benchmarking peers. Our study thereby fills the void in the extant literature that has paid little attention to RPE peers.

We review the related literature in Section II. Section III describes the SEC's new disclosure rules, our sample-selection procedures, and descriptive statistics on RPE use for S&P 1500 firms. In Section IV, we use the implicit approach to evaluate RPE use, where we incorporate explicit RPE contract details. Section V examines firms' decisions to explicitly use RPE in executive compensation contracts, and Section VI investigates the selection of RPE peers. We conclude in Section VII.

II. RELATED LITERATURE

Literature on RPE Use

When common exogenous shocks affect the performance of multiple agents, economic theory predicts that an incentive scheme based on relative performance is superior to an incentive scheme based on individual performance if the agent is sufficiently risk-averse and common uncertainty is high (Lazear and Rosen 1981; Holmstrom 1982; Green and Stokey 1983). Since relative performance reduces the effect of common shocks on performance, RPE allows the risk-averse agent to bear less risk and the principal to better evaluate and motivate the agent's effort. Besides this risk-sharing benefit, tournament theory proposes potential competition benefits of RPE. Under certain circumstances, an incentive scheme based on relative performance can result in higher performance compared to an incentive scheme based on individual performance (Nalebuff and Stiglitz 1983; Hannan et al. 2008).

Despite its theoretical appeal, RPE involves potential costs. Heterogeneity among agents can dilute the benefits of RPE in a standard tournament (Lazear and Rosen 1981). If agents have unequal chances to win the prize given the same level of effort, a tournament can induce disadvantaged agents to shirk (O'Keeffe et al. 1984) and might distort agents' risk choices (e.g., Rosen 1986; Knoeber and Thurman 1994; Hvide 2002). Prior studies also suggest that RPE contracts create adverse incentives for agents to sabotage peer performance, collude with peers, and/or choose inappropriate peer groups (Dye 1984; Gibbons and Murphy 1990; Aggarwal and Samwick 1999; Murphy 2001). Moreover, empirical evidence is consistent with competitive environments and certain executive attributes (such as self-hedging ability and outside employment opportunities) reducing the benefits of RPE in incentive contracting (e.g., Aggarwal and Samwick 1999; Garvey and Milbourn 2003; Rajgopal et al. 2006). The theory and evidence cited above provide potential explanations for a lack of consistent empirical evidence supporting the use of RPE in executive compensation contracts (see Murphy [1999] and Prendergast [1999] for reviews of related literature).

While the majority of empirical studies employ the implicit approach to test RPE use, a few studies have examined *explicit* RPE use in executive compensation contracts (Murphy 1999, 2001; Bannister and Newman 2003; Carter et al. 2009). Murphy (1999) examines 177 large U.S. firms included in the 1997 Towers Perrin survey, and Bannister and Newman (2003) examine proxy disclosures of 160 Fortune 250 firms in fiscal years 1992 and 1993. Both studies provide descriptive statistics about explicit RPE contracts and claim that a lack of empirical support for RPE use could be due to incorrect assumptions and model misspecifications in implicit tests. We investigate

this claim in Section IV. A more recent study by Carter et al. (2009) examines explicit RPE use in performance-vested equity grants for a sample of 129 U.K. firms. Their analysis of firms' choice between RPE use and non-RPE use yields little empirical support for related economic theories.[3] Their U.K. evidence might not generalize to the U.S. setting, however, because performance-vested equity plans are more prevalent in the U.K., mainly due to external pressures from institutional investors (Carter et al. 2009; Ferri 2009).

Literature on RPE Peer Selection

The selection of RPE peers has significant implications for the costs and benefits of RPE, but few theoretical models specifically analyze the selection process. Nevertheless, theories on RPE use provide guidance on how to select RPE peers efficiently. To better shield executives from common exogenous shocks (Lazear and Rosen 1981; Holmstrom 1982; Green and Stokey 1983), selected peers should bear greater common risk with the RPE firms. Moreover, in a tournament, selecting agents with similar ability can reduce potential inefficiencies induced by unequal contests, such as shirking and suboptimal risk choices (e.g., Lazear and Rosen 1981; Hvide 2002). Hence, RPE firms should select peers that exhibit similar ability as themselves.

On the other hand, executive compensation practices are often the product of "managerial power," which allows rent extraction from shareholders (Bebchuk and Fried 2003). In the context of RPE, inefficient peer selection could involve deliberate selection of poorly performing firms as RPE peers to inflate peer-adjusted firm performance and thus allow executives to reap excess compensation (Gibbons and Murphy 1990; Murphy 2001). Another line of research proposes that executive compensation practices reflect symbolic considerations (Westphal and Zajac 1994; Zajac and Westphal 1995), suggesting that RPE firms prefer visible and well-established peers to facilitate the justification of their choices to external constituencies.

Due to limited disclosures about RPE peer groups, prior empirical studies generally assume that RPE peers comprise all firms in the same industry as the RPE firms. Albuquerque (2009) points out that RPE peer selection should consider both industry and size effects, and reports stronger evidence of RPE use when peers are matched on both industry and size, as opposed to industry only. Dikolli et al. (2010) examine the consequences of the implicit test's departure from the details in explicit RPE contracts. They demonstrate analytically that an empiricist's inability to observe the board's actual peer selection and aggregation of peer performance creates measurement errors that can significantly bias against detecting RPE use. Their theoretical proposition is supported by simulation results. These studies highlight the importance of peer-group composition in drawing inferences from traditional implicit tests.

III. INSTITUTIONAL BACKGROUND, SAMPLE, AND DESCRIPTIVE STATISTICS

New Disclosure Rules on Executive Compensation

Amid increased media and shareholder scrutiny of executive pay practices, the SEC amended its rules on executive compensation disclosures in 2006 to increase the transparency of executive compensation contracts. The new disclosure rules were initially proposed in January 2006 and are effective for fiscal years ending on or after December 15, 2006 (see Appendix A for a timeline related to the new disclosure rules). Prior to 2006, proxy disclosures on the details of RPE contracts in the U.S. had been voluntary (Byrd et al. 1998; Carter et al. 2009). Under the new disclosure rules, firms are required to provide a "Compensation Discussion and Analysis" (CD&A) report in their proxy statements, in which firms must provide a detailed description of the

[3] Specifically, Carter et al. (2009) find no evidence that economic determinants of RPE use (such as common risk exposure) affect firms' decisions to use RPE in performance-vested equity grants as predicted by economic theories.

process used to set performance targets and an evaluation of how the performance targets translate into an objective compensation determination.[4] This setting allows us to collect detailed data on explicit RPE contracts for a large sample of U.S. firms.

Mandatory disclosures of RPE use, however, could alter firms' decisions to use RPE and/or the implementation of RPE contracts. To the extent that the new disclosure rules attract public attention concerning the executive pay-setting process, firms could choose to restrain rent-seeking activities and thus potentially improve the compensation-contracting efficiency. Consequently, our findings might understate (overstate) the extent of rent seeking (efficient contracting) in periods prior to this new disclosure regulation.

Sample Selection

We begin with the S&P 1500 firms identified using Compustat's annual file for fiscal year 2006. We retrieve these firms' first proxy statements filed under the new executive compensation disclosure regime.[5] Our initial sample consists of 1,419 firms with available proxy statements.

We determine the use of RPE in executive compensation contracts by reading CD&A reports. Specifically, if the firm states that at least one component of executive compensation (e.g., annual bonus, restricted stock, stock option) is determined based on firm performance relative to a group of peers, the firm is identified as a RPE firm; otherwise, the firm is identified as a non-RPE firm. We take particular care to avoid misclassifying compensation-benchmarking firms as RPE firms. Specifically, if a firm uses peer groups to solely benchmark the *level* of pay without using peer groups to determine the *performance target* when evaluating firm's performance, then the firm is coded as a non-RPE firm. Appendix B provides detailed coding criteria and representative proxy disclosures about RPE contracts in our sample. Our initial sample includes 361 RPE firms and 1,058 non-RPE firms. When firms indicate that RPE is used in setting executive pay, they usually disclose details of the RPE contracts, such as the composition of the RPE peer group and specific performance metrics used for RPE.[6] We collect these detailed RPE contract disclosures as described in the CD&A section.

Finally, we obtain CEO compensation data from ExecuComp, financial data from Compustat, stock return data from CRSP, and analysts' target price forecasts from I/B/E/S.

---

[4] SEC final rule 33-8732a, Item 402(b) (2) (v–vi) states, "In particular, firms need to disclose material information on what specific items of corporate performance are taken into account in setting compensation policies and making compensation decisions… how specific forms of compensation are structured and implemented to reflect these items of the registrant's performance, including whether discretion can be or has been exercised (either to award compensation absent attainment of the relevant performance goal(s) or to reduce or increase the size of any award or payout), identifying any particular exercise of discretion, and stating whether it applied to one or more specified named executive officers or to all compensation subject to the relevant performance goal(s)." Nontransparent or vague disclosures of RPE use may attract SEC attention to be publicized in the SEC's comment letters (see, e.g., the SEC's comment letter on Blackbaud Inc. proxy filings dated October 28, 2008).

[5] Most firms are subject to the new disclosure rules in fiscal year 2006. Since the new rules are effective for fiscal years ending on or after December 15, 2006, and firms are required to file their proxy statements within 120 days after the fiscal year-end, some firms are not subject to the new rules until fiscal year 2007. In our initial sample of 1,419 firms, 37 firms fall into this category, among which 8 firms use RPE. For these firms, we collect information on the use of RPE from fiscal year 2007 proxy statements. For ease of exposition, we use 2006 to refer to the first year under the new disclosure rules throughout the paper.

[6] Firms occasionally comment favorably on the principle of RPE, but do not disclose details on the use of RPE in setting compensation (e.g., no disclosure about the compensation components and performance metrics used for RPE and no details on the RPE peer group; see the last example in Appendix B). We classify these firms as non-RPE firms (66 firms), as under the SEC's new disclosure regime, firms are required to disclose the use and details of RPE contracts if they actually use RPE in setting executive compensation. Our results on firms' decisions to use RPE are qualitatively similar if we exclude these 66 non-RPE firms from the analysis (untabulated).

Descriptive Statistics

Table 1 provides descriptive evidence on the frequency of RPE use and on the main features of RPE contracts. Panel A reports that 361 (about 25 percent) of the sample firms use RPE at least to some extent in determining executive compensation. For comparison, Bannister and Newman (2003) report that 28 percent of their sample firms (which include 160 firms listed on the Fortune 250 index) use RPE over the 1992–1993 period. For S&P 500 firms, which have comparable size to Bannister and Newman's (2003) sample, we find a higher frequency of RPE use (about 38 percent), which reflects increased RPE disclosure under the SEC's new disclosure regime and/or more widespread use of RPE since the early 1990s.[7]

The fraction of RPE firms appears small compared to a theoretical prediction that RPE should be a widespread practice. We offer two explanations. First, we are only able to observe firms pre-committing to a formulaic explicit RPE contract, thereby excluding firms that use RPE implicitly through boards' subjective discretion (e.g., Ferri 2009). Second, as mentioned in Section II, RPE also involves costs, so that firms are apt to use RPE only when the perceived benefits of RPE sufficiently outweigh the costs associated with using RPE.

Table 1, Panel B reports the use of RPE by the type of compensation plans and performance metrics. As can be seen, firms most often apply RPE solely to equity-based compensation plans (214 firms, or 59 percent) and less frequently to cash compensation plans (82 firms, or 23 percent). This finding stands in contrast to the evidence reported by Murphy (1999), based on a survey of 177 large U.S. firms in 1997 that RPE is sometimes used in annual bonus plans but rarely in equity-based compensation plans. Our finding also indicates heavier use of RPE toward equity-based compensation as compared with Bannister and Newman (2003), who report that 44 percent of their sample firms use RPE for long-term plans (including restricted stock, stock options, and long-term incentive plans), based on 160 Fortune 250 firms in the early 1990s. We also find a small percentage of firms (65 firms, or 18 percent) that use RPE in setting both equity-based compensation and cash compensation.

Turning to performance metrics used in RPE contracts, Panel B of Table 1 shows that a majority of RPE firms (206 firms, or 57 percent) rely solely on price-based performance metrics (including stock returns and common shareholder wealth) when implementing RPE. Accounting performance metrics (such as return on equity and earnings per share) are also common in RPE contracts (82 firms, or 23 percent). Price-based RPE metrics are primarily used in equity-based compensation plans (150 firms out of 206 firms, or 73 percent), while cash compensation plans more frequently use accounting-based RPE metrics (35 firms out of 82 firms, or 43 percent). Among the remaining RPE firms, 60 firms (17 percent) use both price-based and accounting performance metrics for RPE purposes. An even smaller percentage of firms (13 firms, or 4 percent) use nonfinancial performance metrics (such as customer satisfaction and market share) in RPE contracts. The infrequent use of nonfinancial metrics potentially reflects difficulties in obtaining peer firms' nonfinancial information (e.g., Ferri 2009). We list commonly used RPE performance metrics in Table 1, Panel C.

Firms that use RPE in setting equity-based compensation can employ RPE in a number of ways. Table 1, Panel D shows that a significant portion of RPE firms (about 43 percent) use RPE

[7] To better understand the effect of the new disclosure rules on firms' disclosures of RPE use, we compare S&P 500 firms' RPE disclosures before and after the new disclosure rules. For the 490 S&P 500 firms in our sample, the percentage of RPE firms in 2005 is about 31 percent (153 firms), compared to 38 percent (185 firms) in 2006. This result appears consistent with more firms adopting RPE in anticipation of more public scrutiny under the new disclosure rules and is also consistent with increased disclosure of RPE use under the new disclosure regime.

TABLE 1

Descriptive Statistics of Relative Performance Evaluation Use in Executive Compensation Plans

Panel A: Relative Performance Evaluation Use[a]

	Number of Firms	Percentage (%)
RPE firms	361	25.44
Non-RPE firms	1,058	74.56
Total	1,419	100.00

Panel B: Relative Performance Evaluation Use by Type of Compensation Plan and Performance Metrics[b]

	Price-Based Metrics	Accounting Metrics	Price-Based and Accounting Metrics	Others	Total
Equity-based compensation plan only					
Restricted stock	126	26	17	3	172
Stock option	3	1	1	0	5
Restricted stock and stock option awards	21	7	6	3	37
Subtotal	150	34	24	6	214
Cash compensation plan only	32	35	12	3	82
Equity-based and cash compensation plans	24	13	24	4	65
Total	206	82	60	13	361

Panel C: Commonly Used Price-Based and Accounting Performance Metrics for Relative Performance Evaluation[c]

	Number of Firms	Percentage (%)
Stock returns	266	73.68
Return on equity	50	13.85
Growth of earnings	42	11.63
Earnings	39	10.81
Growth of sales	24	6.65
Return on asset	14	3.88
Cash flow	13	3.60

Panel D: Details of Relative Performance Evaluation Use in Equity-Based Compensation Plan[d]

	Number of Firms	Percentage (%)
RPE determines		
Vesting condition of performance share units (type 1)	120	43.01
Size of restricted stocks and stock options (type 2)	94	33.69
Vesting condition of restricted stocks and stock options (type 3)	61	21.86
Both type 1 and type 3	2	0.72
Both type 2 and type 3	2	0.72
Total	279	100.00

(continued on next page)

Panel E: Relative Performance Evaluation Use by Type of Compensation Plan and Peer Group Choice[e]

	Self-Selected Peers	Market/Industry Index	Market/Industry Index and Self-Selected Peers	Total
Equity-based compensation plan only				
Restricted stock	91	72	9	172
Stock option	4	1	0	5
Restricted stock and stock option awards	21	10	6	37
Subtotal	116	83	15	214
Cash compensation plan only	51	25	6	82
Equity-based and cash compensation plans	42	19	4	65
Total	209	127	25	361

Panel F: Number of Peers in the Self-Selected Peer Groups and Industry Similarity and S&P 1500 Identity for Self-Selected Peers[f]

	Mean	Standard Deviation	25%	Median	75%	n
Number of self-selected peers	14.694	8.721	9	13	18	232
Percentage of self-selected peers in the same two-digit SIC industry as the RPE firms (%)	60.92	33.39	34.85	66.67	91.67	232
Percentage of self-selected peers in the same three-digit SIC industry as the RPE firms (%)	47.92	34.93	14.84	45.45	83.33	232
Percentage of self-selected peers in the S&P 1500 index (%)	70.37	22.42	58.58	74.60	86.36	232
Percentage of self-selected peers in the same S&P 1500 sub-index as the RPE firms (%)	45.11	27.77	23.67	40.00	67.71	232

[a] RPE (Non-RPE) firms include 361 (1,058) S&P 1500 firms listed in the Compustat annual file for 2006 that explicitly disclose (do not mention) the use of relative performance evaluation in executive compensation plans in their first annual proxies filed under the SEC's 2006 executive compensation disclosure rules. We are unable to find proxy statements for 81 S&P 1500 firms.

[b] "Equity-based compensation plan only" category includes firms that use relative performance evaluation in setting equity-based compensation (such as restricted stock and stock option). "Cash compensation plan only" category includes firms that use relative performance evaluation in setting cash compensation (such as annual bonus and long-term incentive payout). "Equity-based and cash compensation plans" category includes firms that use RPE in both equity-based and cash compensation plans.

"Price-Based Metrics" category includes firms that use price-based performance metrics (such as stock returns and shareholder wealth) to implement relative performance evaluation in compensation plans. "Accounting Metrics" category includes firms that use accounting performance metrics (such as return-on-equity, earnings per share, earnings growth) to implement relative performance evaluation in compensation plans. "Price-Based and Accounting Metrics" category includes firms that use both price-based and accounting performance metrics to implement relative performance evaluation in compensation plans. "Others" category includes firms that use nonfinancial performance metrics (such as customer satisfaction and market share) to implement relative performance evaluation in compensation plans. See Panel C for a more detailed breakdown of performance metrics.

[c] The percentages sum to more than 100 percent as some firms employ multiple performance metrics when applying relative performance evaluation in compensation plans.

[d] "Vesting condition of performance share units" category includes firms that use RPE to determine the vesting of "performance share units (PSU)." Performance shares are similar to restricted stocks in that they fully vest at the end of a certain performance period starting with the grant year (usually three years). However, the range of vesting amount for PSU is usually between 0 percent and 200 percent (sometimes 150 percent), which is different from that for restricted stock units (RSU) capped at 100 percent. "Size of restricted stocks and stock options" category includes firms that use RPE to determine the size of granted restricted stocks and stock options. "Vesting condition of restricted stocks and

(continued on next page)

stock options" category includes firms whose stock options or RSUs are vested conditioned on achievement of RPE goals. This category also includes firms for which the vesting schedule of stock options or restricted stock is accelerated if the company achieves a certain RPE target.

[e] "Self-Selected Peers" category includes firms that chose individual firms as the peer groups to implement relative performance evaluation in compensation plans. "Market/Industry Index" category includes firms that chose published market/industry indices as the peer groups to implement relative performance evaluation in compensation plans. "Market/Industry Index and Self-Selected Peers" includes firms that chose both published market/industry indices and individual firms as the peer groups to implement relative performance evaluation in compensation.

[f] Two firms mentioned the use of self-selected peers in implementing relative performance evaluation in executive compensation plans, but did not disclose the composition of the RPE peer group, which reduces the sample size to 232 firms. On a firm level, there are 39 (25) RPE firms whose self-selected peer firms all come from the same two-digit (three-digit) SIC industry; there are 25 (10) RPE firms whose self-selected peer firms all come from the S&P 1500 index (the same S&P 1500 sub-index) as the RPE firms.

to determine the vesting of performance share units.[8] A number of firms (about 34 percent) use RPE to determine the size of restricted stock and stock option grants. Other firms (about 22 percent) link the achievement of RPE performance target to vesting conditions of restricted stocks and stock options. These statistics indicate that the use of RPE in equity grants now commonly involves vesting conditions among U.S. firms.

Finally, we examine the choice and composition of RPE peer groups. Table 1, Panel E reports that a majority of RPE firms (209 firms, or 58 percent) use self-selected RPE peer groups, and about one-third of RPE firms (127 firms, or 35 percent) employ published market or industry indices. A small number of RPE firms (25 firms, or 7 percent) use both published indices and self-selected peer groups in RPE contracts. The tendency to use self-selected peer groups (as opposed to published indices) does not seem to differ notably across equity-based compensation plans and cash compensation plans.

Table 1, Panel F shows that the mean (median) number of peers in self-selected peer groups is about 15 (13) firms, in contrast to a mean (median) of 60 (65) firms in these firms' corresponding two-digit SIC industry. Moreover, only 61 percent (48 percent) of self-selected peers operate in the same two-digit (three-digit) SIC industry as the RPE firms. At the firm level, there are only 39 (25) RPE firms whose self-selected peers all come from the same two-digit (three-digit) SIC industry. These statistics reinforce the concern that the key assumption underlying implicit tests—that the RPE peer group consists of all firms from the same industry—is inaccurate.[9] Finally, about 70 percent of self-selected peers belong to the S&P 1500 index, and about 45 percent share the same S&P sub-index (S&P 500, Mid-Cap 400, and Small-Cap 600) as the RPE firms. At the firm level, there are 25 (10) RPE firms whose self-selected peers all come from the S&P 1500 index (the same S&P 1500 sub-index).

Table 2 provides descriptive statistics on firm characteristics, management attributes, and corporate governance for the sample firms. As shown, firms that disclose RPE use in executive compensation contracts differ from non-RPE firms along many dimensions, including competitive

[8] Performance shares are similar to restricted stocks in that they fully vest at the end of a specified performance period, starting with the grant year (usually three years). However, the vesting amount for performance shares ranges between 0 percent and 200 percent (see the example of PG&E Corp in Appendix B), whereas that for restricted stocks is capped at 100 percent.

[9] A possible reason for a RPE firm to select peers outside its (primary) industry is that the RPE firm follows diversified business strategies and operates with multiple segments, which requires a diversified peer group to appropriately benchmark firm performance. We find that, of the 232 RPE firms that have self-selected peer groups, 157 firms (about 68 percent) operate with multiple business segments. Multiple-segment RPE firms have roughly 61 percent (45 percent) of peers selected from the same two-digit (three-digit) SIC industry, while single-segment RPE firms have about 72 percent (65 percent) of peers selected from the same two-digit (three-digit) SIC industry.

TABLE 2

Descriptive Statistics of the Sample

	Non-RPE Firms (n = 1,058)		RPE Firms (n = 361)			
	Mean	**Median**	**Mean**	**Median**	**Mean Differences**	**Median Differences**
Firm Characteristics						
Common_Risk	0.285	0.257	0.384	0.370	−0.099***	−0.113***
|Size_Rkadj|	5.348	5.081	6.325	5.969	−0.977***	−0.888***
|Diversity_Rkadj|	0.019	0.007	0.024	0.015	−0.005***	−0.008***
|Return_Rkadj|	0.049	0.024	0.040	0.022	0.009**	0.002
Industry_Concentration	0.124	0.085	0.110	0.072	0.014**	0.013***
BM	0.600	0.602	0.682	0.692	−0.082***	−0.090***
Size	7.654	7.466	8.484	8.355	−0.830***	−0.889***
ROA_Indadj	0.014	0.000	0.005	0.000	0.009	0.000
Return_Indadj	0.017	−0.007	0.019	−0.001	−0.002	−0.006
CEO Attributes						
CEO_Wealth	8.972	9.868	9.435	10.011	−0.463	−0.143
CEO_Age	55.882	56.000	56.053	56.000	−0.171	0.000
Corporate Governance						
Top5_Instown	0.392	0.372	0.374	0.353	0.018**	0.019**
Activist_Instown	0.028	0.028	0.028	0.029	0.000	−0.001**
CEO/Chair	0.601	1.000	0.659	1.000	−0.058*	0.000
Board_Independence	0.817	0.857	0.857	0.889	−0.040***	−0.032***
Board_Size	8.957	9.000	10.518	10.000	−1.561***	−1.000***
CompConsultant	0.798	1.000	0.904	1.000	−0.106***	0.000

*, **, *** Indicate significance at less than the 10 percent, 5 percent, and 1 percent levels, respectively, based on two-tailed t-tests (z-tests) on mean (median) differences.

The sample includes S&P 1500 firms listed in the Compustat annual file for 2006 that have available proxy statements from the SEC's EDGAR database. RPE (Non-RPE) firms include 361 (1,058) S&P 1500 firms that explicitly disclose (do not mention) the use of relative performance evaluation in executive compensation plans in their first annual proxies filed under the SEC's 2006 executive compensation disclosure rules.

Variables are measured over or at the end of fiscal year 2005 (2006) if a firm's first proxy filed under the SEC's 2006 executive compensation disclosure rules is for fiscal year 2006 (2007). Firm characteristics and CEO attributes (except *Industry_Concentration* and *CEO_Age*) are winsorized at the top and bottom one percentiles. Number of observations varies depending on data availability.

Variable Definitions:

Common_Risk = proportion of firm-level stock return variance that is explained by value-weighted industry stock returns, measured by the R^2 from regressing firm-level stock returns on value-weighted industry stock returns over the prior 36 months;

|Size_Rkadj| = natural logarithm of the absolute difference between the firm's market value of equity and the median market value of equity for the firm's corresponding decile;

|Diversity_Rkadj| = natural logarithm of the absolute difference between the firm's segment concentration and the median segment concentration for the firm's corresponding decile. Segment concentration is the sum of the squares of each segment's sales as a percentage of the total firm sales;

|Return_Rkadj| = absolute difference between the firm's annual stock returns and the median annual stock returns for the firm's corresponding decile;

Industry_Concentration = sum of the squares of the market shares of the firms' sales within each two-digit SIC industry;

BM = book value of assets divided by the sum of the market value of equity and book value of liabilities;

Size = natural logarithm of the market value of equity;

(continued on next page)

TABLE 2 (continued)

ROA_Indadj = return-on-assets minus the median return-on-assets for the same industry (two-digit SIC code);
Return_Indadj = buy-and-hold annual stock returns minus the median buy-and-hold annual stock returns for the same industry (two-digit SIC code);
CEO_Wealth = natural logarithm of the value of equity (including both stocks and stock options) held by the CEO;
CEO_Age = age of the CEO;
CompConsultant = 1 if the firm uses compensation consultant in the pay-setting process (as disclosed in their first annual proxies filed under the SEC's 2006 executive compensation disclosure rules), and 0 otherwise;
Top5_Instown = stock ownership by the top five institutions as a percentage of total institutional ownership;
Activist_Instown = percentage of holdings by activist institutions as defined by Cremers and Nair (2005);
CEO/Chair = 1 if the CEO serves as Chairman of the Board, and 0 otherwise;
Board_Independence = percentage of independent directors serving on the board; and
Board_Size = number of directors serving on the board.

environment, growth opportunities, common risk exposure, and market capitalization. In addition, RPE firms have different ownership structure and board attributes than non-RPE firms, and more frequently employ compensation consultants when setting executive compensation. We consider these differences when conducting multivariate analyses in Section V.

IV. INCORPORATING EXPLICIT RPE CONTRACT DETAILS IN IMPLICIT TESTS

Prior empirical studies generally employ an implicit approach to test for the prevalence of RPE use *without* referring to explicit RPE contracts. In this section, we examine the extent to which incorporating details of explicit RPE contracts affects inferences drawn under the implicit approach. Our approach, while incorporating important features of RPE contracts, still relies on assumptions regarding how RPE is actually utilized in setting executive pay.[10] Nevertheless, such evidence would highlight limitations of implicit tests and underscore the importance of considering explicit RPE contract details in drawing conclusions about RPE use.

We start by following Albuquerque (2009)'s implicit approach with one exception. Albuquerque (2009) included CEO fixed effects for her multiple-year sample period of 1992–2005. We estimate the following equation using S&P 1500 firms, where we limit the analysis to 2006 because we only collect information about explicit RPE contracts for 2006:

$$CEOPay_t = \alpha_0 + \alpha_1 FirmPerf_t + \alpha_2 PeerPerf_t + \alpha_3 ControlVariables_t + \varepsilon_t. \quad (1)$$

In Equation (1), the dependent variable, *CEOPay*, is CEO total compensation, measured as the sum of salary, bonus, the fair value of stock option awards and restricted stock awards (per SFAS 123R), the change in deferred compensation, non-equity incentive plan compensation, and other compensation, in thousands of dollars (i.e., TDC1 from ExecuComp). To mitigate skewness in CEO total compensation, we follow prior studies and use its natural logarithm in the estimation (e.g., Murphy 1999; Albuquerque 2009). The independent variables, $FirmPerf_t$ and $PeerPerf_t$, are 12-month buy-and-hold stock returns during the fiscal year for the firm and its RPE peer group, respectively. Albuquerque (2009) proposes firm size as an important factor in RPE peer selection, while prior implicit tests utilize misspecified peer groups by combining same-industry firms with-

[10] For instance, we assume that all firms use stock return as the RPE-based performance metric and the median performance among peers determines the hurdle rate in achieving RPE-based performance targets.

out considering firm size. We follow Albuquerque's (2009) recommendation and measure $PeerPerf_t$ based on the median stock return for firms in the same size quartile and same industry. If firms filter out peer performance in setting CEO pay, then we expect to find $\alpha_2 < 0$.

Equation (1) also controls for firm characteristics and governance attributes that can influence CEO pay, including firm size, growth options, CEO tenure, regulated industry membership, idiosyncratic variance, CEO/Chair duality, CEO stock ownership, and whether the CEO is involved in an interlocking relationship. To account for potential outliers, we winsorize non-dichotomous variables at top and bottom one percentiles (except that idiosyncratic variance is winsorized at the top one percentile only). We further include industry indicator variables to account for variations in CEO pay at the industry level.

Table 3 presents the results from estimating Equation (1). The first column reports results for our full sample of S&P 1500 firms. The sample size is smaller than that reported in Table 1, Panel A (reduced by 291 firms) due to (1) additional sample restrictions (excluding new CEOs who take the position in the middle of 2006 and excluding firms reporting negative total compensation), and (2) missing values on control variables. As shown, the coefficient on industry-size-matched peer returns is insignificant, inconsistent with RPE use.[11] The lack of evidence in support of RPE use could result from infrequent use of RPE in 2006, as only 25 percent of our sample firms explicitly disclose the use of RPE in setting executive pay. If the infrequent use of RPE underlies the insignificant result above, then we expect to observe stronger evidence on RPE use for RPE firms than for non-RPE firms. Columns (2) and (3) report the results for non-RPE firms and RPE firms, respectively. We find that for both non-RPE firms and RPE firms, the coefficient on industry-size matched peer returns is insignificant. Hence, even among firms explicitly disclosing RPE use, the implicit test provides no evidence that these firms filter out peer performance when setting CEO pay. These results indicate that the infrequent use of RPE is unlikely to be the dominant reason for the insignificant evidence on RPE use in the implicit test.

To further refine the implicit test, we consider disclosed RPE contract details. First, we incorporate RPE peer-group composition by measuring $PeerPerf_t$ based on the median stock return of self-selected RPE peers as disclosed in firms' 2006 proxy statements. Table 3, Column (4) shows that the coefficient on self-selected peer returns is significantly negative, consistent with firms using these self-selected peers for RPE purposes when setting CEO pay. The contrasting results across Columns (3) and (4) highlight the importance of peer-group composition in testing for RPE use.[12] The implicit test is likely to produce misleading results due to inaccurate identification of RPE peers employed in the pay-setting process.

Second, we consider performance-based vesting, a design feature of RPE contracts that has gained popularity in the United States in recent years (Gerakos et al. 2007). The traditional implicit tests use *current* performance for RPE firms and their peers to explain executive pay. However, as Table 1, Panel D documents, firms often link RPE-based performance targets to vesting conditions of equity grants, so that *future* performance determines the value of equity grants (Gerakos et al. 2007; Carter et al. 2009). Under SFAS 123R, firms are required to report the fair value of performance-vested equity grants based on estimates of future performance for both

[11] This result appears to be inconsistent with Albuquerque's (2009) main finding of a negative coefficient on industry-size-matched peer returns. This inconsistency is driven by our different sample period: Albuquerque's (2009) sample period ranges from 1992 to 2005, while our sample period is limited to 2006.

[12] The sample size under Column (3) is larger than the sample size under Column (4) because Column (4) focuses attention on RPE firms disclosing self-selected RPE peers. Results are qualitatively similar if we restrict to RPE firms disclosing self-selected RPE peers in Column (3). The sample size under Column (4) is smaller than the sample size reported in Table 1, Panel E (i.e., 209 firms with self-selected RPE peer groups) because of additional sample restrictions (excluding new CEOs who take the position in the middle of 2006 and excluding firms reporting negative total compensation) and missing values on control variables.

TABLE 3

Implicit Test of RPE Use With and Without Incorporating Explicit RPE Contract Details

		Without Incorporating Explicit RPE Contract Details			Incorporating Explicit RPE Contract Details		
					RPE Firms with Self-Selected Peers		
		Full Sample (S&P 1500)	Non-RPE Firms	RPE Firms	All	Linked to *Current* Peer Performance	Linked to *Future* Peer Performance
		(1)	(2)	(3)	(4)	(5)	(6)
	Predicted Sign	Coefficient (p-value)	Coefficient (p-value)	Coefficient (p-value)	Coefficient (p-value)	Coefficient (p-value)	Coefficient (p-value)
Independent Variables							
Firm Stock Return	+	0.329	0.273	0.928	0.990	0.759	0.694
		(0.001)	(0.013)	(<0.001)	(<0.001)	(0.077)	(0.039)
Peer Stock Return (Industry/Size)	−	0.284	0.053	0.604			
		(0.413)	(0.898)	(0.346)			
Peer Stock Return (Self-Selected Peers)	−				−0.995	−1.671	0.258
					(0.034)	(0.029)	(0.849)
Control Variables							
Firm Size (Sales)	+	0.468	0.478	0.427	0.454	0.450	0.387
		(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)
Growth Option	+	0.080	0.065	0.225	0.261	0.241	0.323
		(0.005)	(0.033)	(<0.001)	(<0.001)	(0.006)	(<0.001)
CEO Tenure	+	−0.074	−0.056	−0.136	−0.075	−0.074	−0.237
		(0.040)	(0.182)	(0.056)	(0.392)	(0.642)	(0.141)
Regulation Industry	−	−0.245	−0.204	−0.472	−0.752	−1.042	−0.790
		(0.097)	(0.421)	(0.379)	(0.002)	(0.018)	(0.069)
Idiosyncratic Variance	+	0.871	0.427	3.890	2.050	2.509	−1.803
		(0.411)	(0.720)	(0.105)	(0.348)	(0.566)	(0.746)

(continued on next page)

TABLE 3 (continued)

		Without Incorporating Explicit RPE Contract Details			Incorporating Explicit RPE Contract Details		
					RPE Firms with Self-Selected Peers		
		Full Sample (S&P 1500)	Non-RPE Firms	RPE Firms	All	Linked to *Current* Peer Performance	Linked to *Future* Peer Performance
		(1)	(2)	(3)	(4)	(5)	(6)
	Predicted Sign	Coefficient (p-value)	Coefficient (p-value)	Coefficient (p-value)	Coefficient (p-value)	Coefficient (p-value)	Coefficient (p-value)
CEO/Chair	+	0.233	0.201	0.375	0.468	0.557	0.578
		(<0.001)	(0.001)	(<0.001)	(<0.001)	(0.006)	(0.003)
CEO Ownership	+	−0.068	−0.073	0.061	0.154	0.104	0.095
		(0.292)	(0.315)	(0.709)	(0.461)	(0.688)	(0.813)
Interlock	+	−0.938	−1.436	0.585	0.403	0.413	NA
		(0.026)	(<0.001)	(0.633)	(0.707)	(0.713)	
Intercept		4.446	4.714	4.231	4.477	4.971	5.009
		(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)
Industry fixed effects		Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2		0.550	0.536	0.659	0.728	0.741	0.805
Number of observations		1,128	837	291	184	95	89

The full sample includes 1,128 S&P 1500 firms from the ExecuComp database for the year 2006 merged with Compustat and CRSP databases. The standard errors are heteroscedasticity-consistent using the Huber-White correction and are clustered by firm. The reported p-values in parentheses are two-tailed. The dependent variable, *CEOPay*, is the natural logarithm of total CEO compensation for 2006 in thousands of dollars (TDC1).

Variable Definitions:
Firm Stock Return = 12-month buy-and-hold stock returns for 2006;
Peer Stock Return = measured in two ways: *Peer Stock Return* (*Industry/Size*) is the median 12-month buy-and-hold stock returns for 2006 (excluding own-firm performance)

(continued on next page)

TABLE 3 (continued)

for the industry and size-matched peer group (see Albuquerque (2009) for details); *Peer Stock Return* (*Self-Selected Peers*) is the median 12-month buy-and-hold stock returns for 2006 for the self-selected peer group by RPE firms;

Firm Size (Sales) = natural logarithm of sales for 2005;

Growth Option = ratio of the market value of equity to the book value of assets at the beginning of 2006;

CEO Tenure = natural logarithm of the number of years since the CEO assumed office;

Regulation Industry = an indicator variable equal to 1 for firms in the gas and electric industries with SIC codes from 4900 to 4939, and 0 otherwise;

Idiosyncratic Variance = difference between firm-level stock return variance and the industry's average return variance, calculated over the previous 36 months;

CEO/Chair = indicator variable equal to 1 if the CEO is the board chair, and 0 otherwise;

CEO Ownership = indicator variable equal to 1 if the proportion of shares held by the CEO is lower than the sample median, and 0 otherwise; and

Interlock = indicator variable equal to 1 if the CEO is involved in an interlock relationship requiring disclosure in the proxy statement, and 0 otherwise.

the RPE firms and their peers, as well as the associated likelihood of future vesting realizations. Because compensation committees' projections of future performance can differ from current performance, the use of current performance in testing for RPE use could produce low-power tests when RPE-based performance targets are linked to future performance rather than current performance. The last two columns of Table 3 examine this issue. As shown in Column (5), when RPE use is linked to current peer performance, the coefficient on self-selected peer returns is significantly negative, supporting the use of RPE. In contrast, Column (6) shows that, when RPE use is linked to future peer performance (through equity grants' vesting conditions), there is no evidence of RPE use. Hence, failure to consider this feature of RPE contracts is likely to confound empirical results based on implicit tests of RPE use.[13]

V. USE OF RPE IN EXECUTIVE COMPENSATION CONTRACTS

Prior studies have proposed a variety of factors that affect firms' decisions to use RPE in executive compensation contracts. However, these studies rely on the implicit approach and mostly examine one contextual factor at a time without controlling for alternative factors. In this section, we employ explicit disclosures on RPE use to simultaneously examine multiple factors that influence the decision to incorporate RPE into executive compensation contracts.

Determinants of RPE Use

Common Risk

Theory predicts that the risk-sharing benefit of RPE increases in the extent to which common risk affects both firm and peer performance (e.g., Holmstrom 1982).[14] We thus expect firms to be more likely to use RPE when their performance is more heavily influenced by common risk. We measure common risk based on the proportion of the variance of firm-level stock returns that can be explained by industry returns, defined as the R^2 from regressing firm-level stock returns on value-weighted industry stock returns over the prior 36 months (*Common_Risk*).

Availability of Similar Peers

Tournament theory implies that including firms with unequal ability in the RPE peer group could induce executives to shirk or to pursue inefficient investments. Firms are thus less likely to use RPE if potential peers all exhibit dissimilar ability. We argue that firm size and industry diversification reflect organizational complexity, and firm performance reflects existing economic conditions, all of which affect executives' ability to compete with potential peers. To measure the availability of peers with similar ability, we first divide firms within the same two-digit SIC industry into deciles based on market value of equity, segment concentration defined as the Herfindahl-Hirschman Index of segment sales (Rose and Shepard 1997; Bushman et al. 2004), or annual stock returns. The availability of similar peers is then measured by the absolute differences in market value of equity, segment concentration, and annual stock returns, respectively, between the firm and the median of the decile to which the firm belongs (|*Size_Rkadj*|, |*Diversity_Rkadj*|, and |*Return_Rkadj*|, respectively).

---

[13] Other features of RPE contracts relate to compensation components, performance metrics, hurdle rates in achieving RPE-based performance targets (e.g., median, 25 percent, or 75 percent of peer group performance), etc. Combining *all* of the features of RPE contracts in studying RPE use would substantially reduce our sample size and, hence, is not feasible to implement.

[14] While it is desirable to filter out common shocks for better risk-sharing, some firms may deliberately choose not to use RPE to motivate executives to manage risks resulting from such exogenous shocks (e.g., hedging in oil and gas industry).

Industry Competition

Aggarwal and Samwick (1999) and Joh (1999) argue that firms facing a more competitive environment are less likely to use RPE due to the concern that RPE may encourage destructive competition. Following these studies, we use industry concentration, measured as the Herfindahl-Hirschman Index of sales within each two-digit SIC industry, as an inverse proxy for industry competition (*Industry_Concentration*).

Growth Opportunities

Murphy (2001) predicts that firms with high-growth opportunities are more likely to adopt external standards (such as peer performance) in setting executive pay because internal standards (such as historical earnings growth) provide managers of high-growth firms stronger incentives to smooth performance in response to fears of budget ratcheting (Leone and Rock 2002). We use the book-to-market ratio (*BM*) as an inverse proxy for growth opportunities.

Size

Himmelberg and Hubbard (2000) use firm size to capture CEO talent. They propose that the reservation wage for talented CEOs covaries with the state of the economy, such that firms are less likely to filter out industry and market-wide performance when compensating talented CEOs (also see Rajgopal et al. 2006). Alternatively, firm size could serve as a crude proxy for public scrutiny and shareholder concerns about executive pay practices (Bannister and Newman 2003), which may induce firms to commit to explicit RPE use. We use the market value of equity to measure firm size (*Size*).

Firm Performance

Firm performance could also proxy for CEO talent and, hence, could be negatively related with RPE use (Himmelberg and Hubbard 2000). Alternatively, firms having superior performance could prefer RPE as a convenient tool to justify higher executive pay. We include industry-adjusted operating performance (*ROA_Indadj*) and stock performance (*Return_Indadj*) to capture performance effects on RPE use.

Executive Attributes

Garvey and Milbourn (2003) argue that the theoretical benefit of RPE is diminished to the extent that agents can hedge market risk. Hence, firms are less likely to use RPE when their CEOs display greater self-hedging ability. Following Garvey and Milbourn (2003), we use the value of CEO equity holdings (*CEO_Wealth*) and CEO age (*CEO_Age*) to proxy for CEOs' self-hedging ability.

Ownership Structure and Board Attributes

Bertrand and Mullainathan (2001) document that CEO pay is positively associated with exogenous shocks and that such "pay-for-luck" is less pronounced in better-governed firms. Garvey and Milbourn (2006) further show that executive pay is more sensitive to good luck than to bad luck and that this asymmetry is more pronounced for firms with weaker governance. Since RPE can limit the pay-for-luck practice and also has risk-sharing benefit, we expect that RPE use is positively associated with the quality of corporate governance. We measure governance quality using institutional ownership concentration (*Top5_Instown*), activist institutional ownership (*Activist_Instown*) as defined in Cremers and Nair (2005), CEO/Chair duality (*CEO/Chair*), the proportion of outside directors on the board (*Board_Independence*), and the number of directors sitting on the board (*Board_Size*).

Compensation Consultant

Compensation consultants have access to proprietary information about industry-wide compensation practices and potential competitors (Cadman et al. 2010; Murphy and Sandino 2010). Since the design and implementation of RPE contracts demand specialized knowledge and expertise, the availability of professional advice from compensation consultants is likely to influence firms' decisions to use RPE. We construct an indicator variable to identify firms that use compensation consultants in setting executive pay (*CompConsultant*), based on firms' first proxy disclosures under the SEC's new disclosure rules.

Results on RPE Use

To examine firms' decisions to use RPE in executive compensation contracts, we estimate the following multivariate logistic regression:

$$\begin{aligned}Prob(RPE_t = 1) = \Phi(&\alpha_0 + \alpha_1 Common\_Risk_{t-1} + \alpha_2 |Size\_Rkadj|_{t-1} + \alpha_3 |Diversity\_Rkadj|_{t-1}\\ &+ \alpha_4 |Return\_Rkadj|_{t-1} + \alpha_5 Industry\_Concentration_{t-1} + \alpha_6 BM_{t-1}\\ &+ \alpha_7 Size_{t-1} + \alpha_8 ROA\_Indadj_{t-1} + \alpha_9 Return\_Indadj_{t-1} + \alpha_{10} CEO\_Wealth_{t-1}\\ &+ \alpha_{11} CEO\_Age_{t-1} + \alpha_{12} Top5\_Instown_{t-1} + \alpha_{13} Activist\_Instown_{t-1}\\ &+ \alpha_{14} CEO/Chair_{t-1} + \alpha_{15} Board\_Independence_{t-1} + \alpha_{16} Board\_Size_{t-1}\\ &+ \alpha_{17} CompConsultant_t + \varepsilon_t).\end{aligned} \tag{2}$$

The dependent variable, *RPE*, is an indicator variable that equals 1 for RPE firms and 0 for non-RPE firms. The independent variables include firm characteristics, CEO attributes, and corporate governance factors that potentially influence the use of RPE.

Table 4 reports the results from estimating Equation (2). We find that firms exposed to greater common risk are more likely to use RPE, in line with the notion that RPE is more effective at removing exogenous shocks in performance evaluation when the firm shares greater common risk with its peers (Janakiraman et al. 1992). In contrast to the prediction that more intense product market competition (lower industry concentration) discourages the use of RPE (Aggarwal and Samwick 1999), we find that firms from less concentrated industries are more likely to implement RPE in compensation plans. This result, however, is consistent with the evidence in DeFond and Park (1999) that a more competitive environment is characterized by a higher degree of common risk and, hence, brings greater benefits to those who use RPE.[15] We also find that firms exhibiting fewer growth opportunities (higher book-to-market) are more likely to use RPE, which contradicts Murphy's (2001) prediction but is in line with the argument in Albuquerque (2007) that RPE is less beneficial for firms with high-growth options, since peer performance is a less informative signal of external shocks for these firms.[16]

Furthermore, we find that larger firms are more likely to use RPE, consistent with firm size serving as a proxy for public scrutiny and shareholder concerns about executive pay. We also find that firms are less likely to use RPE when the value of CEO equity holdings is large, consistent

---

[15] DeFond and Park (1999) examine RPE use in management turnover decisions and find stronger evidence of RPE use in highly competitive industries than in less competitive industries. Our finding that *Industry_Concentration* (inverse proxy for industry competition) is negatively related to RPE use, even after controlling for *Common_Risk*, suggests that industry concentration may capture common risk above and beyond *Common_Risk*.

[16] Albuquerque (2007) argues that growth options affect a firm's common risk exposure and, hence, the informativeness of peer performance about the firm's common risk. Given we have controlled for *Common_Risk*, the positive relation between book-to-market and RPE use suggests that book-to-market may capture common risk above and beyond *Common_Risk*.

TABLE 4

Logistic Regression of Relative Performance Evaluation Use in Executive Compensation Contracts

Independent Variables	Predicted Sign	Coefficient (p-value)	Change in Probability Q1 versus Q3 Values
Firm Characteristics			
Common_Risk	+	1.206 (0.016)	0.088
|*Size_Rkadj*|	−	−0.087 (0.231)	−0.049
|*Diversity_Rkadj*|	+/−	4.712 (0.222)	0.039
|*Return_Rkadj*|	−	1.257 (0.303)	0.010
Industry_Concentration	+	−6.022 (0.001)	−0.097
BM	−	1.499 (0.005)	0.118
Size	+/−	0.451 (0.001)	0.026
ROA_Indadj	+/−	−0.423 (0.641)	−0.009
Return_Indadj	+/−	−0.016 (0.874)	−0.001
CEO Attributes			
CEO_Wealth	−	−0.168 (0.002)	−0.051
CEO_Age	−	0.006 (0.664)	0.014
Corporate Governance			
Top5_Instown	+	−0.321 (0.738)	−0.010
Activist_Instown	+	4.837 (0.688)	0.011
CEO/Chair	−	0.193 (0.322)	0.048
Board_Independence	+	4.614 (0.000)	0.011
Board_Size	+/−	0.166 (0.001)	0.071
CompConsultant	+	0.328 (0.052)	0.081
Intercept		−9.861 (<0.001)	
Number of RPE/Non-RPE firms		255/640	
Percent concordant/discordant		76.9/22.9	
Pseudo R^2		0.255	
Wald χ^2 (p-value)		118.071 (<0.001)	

(continued on next page)

TABLE 4 (continued)

The sample includes 895 S&P 1500 firms listed in the Compustat annual file for 2006 that have available information on relative performance evaluation use in executive compensation contracts and available information on regression variables. The dependent variable is *RPE*, which is equal to 1 if the firm explicitly reported RPE use in executive compensation plans in its first annual proxy filed under the SEC's 2006 executive compensation disclosure rules, and 0 otherwise. See Table 2 for other variable definitions. Firm characteristics and CEO attributes (except *Industry_Concentration* and *CEO_Age*) are winsorized at the top and bottom one percentiles. The "Change in Probability" column shows the change in the probability of using RPE as the result of moving from the first to the third quartile value of the variable of interest (for indicator variables, moving from 0 to 1), holding all other variables constant at their mean values (for indicator variables, the benchmark probability is determined with the 0 value). The reported p-values in parentheses are two-tailed.

with Garvey and Milbourn (2003). Finally, firms with more independent boards are more likely to use RPE, consistent with stronger internal governance favoring the use of RPE, and firms that hire compensation consultants are also more likely to use RPE, suggesting that compensation consultants' specialized knowledge and expertise facilitate the use of RPE. Among all determinants, growth opportunities, common risk exposure, industry concentration, compensation consultants, and board size have a relatively large impact on RPE use.[17]

In summary, our evidence confirms the importance of common risk and CEOs' self-hedging ability in firms' decisions to use RPE, and it shows that firms with more independent boards and firms employing compensation consultants are more likely to use RPE. However, some of our results contradict prior evidence based on implicit tests (such as the effect of industry competition). We also find little evidence that the availability of similar peers affects firms' decision to use RPE. Taken together, the evidence suggests that firms consider both costs and benefits when deciding to use RPE in executive compensation contracts. A potential caveat is that firms could have adopted RPE contracts prior to the new disclosure regulation, and, hence, our causal inferences should be interpreted with caution.

We acknowledge that firms can use RPE implicitly through boards' subjective discretion, rather than pre-committing to a formulaic explicit RPE contract (Ferri 2009). We also note that the explanatory variables in Equation (2) likely capture factors that influence a firm's decision to use explicit versus implicit RPE. From the efficient-contracting perspective, when RPE can bring greater net benefits (e.g., higher common risk exposure), it is likely optimal for the principal and the agent to pre-commit to an explicit RPE contract. From the rent-seeking perspective, to the extent that more flexibility in the pay-setting process allows managers to reap greater private benefits, firms with weaker corporate governance (e.g., less independent boards) are less likely to commit to an explicit RPE contract. Furthermore, firms facing intense public scrutiny for transparent compensation disclosures (e.g., larger firms) are more likely to explicitly commit to RPE to cater to shareholder preference by reducing the extent of board discretion. We refrain, however, from drawing inferences on explicit versus implicit RPE use, as a thorough investigation of this topic is beyond the scope of our study.

[17] Firms may use RPE to determine anywhere from a substantial portion to a minimal amount of executive pay. While it is difficult to ascertain the amount of executive pay that directly results from RPE use, we develop a rough proxy for the extent of RPE use (*RPE_prop)* based on the ratio of RPE-involved compensation (e.g., bonus and long-term incentive payouts, restricted stocks, or stock options) to total compensation for CEOs. We exclude RPE firms with *RPE_prop* less than 10 percent, and re-estimate Equation (2). The untabulated results are qualitatively similar as those reported. We also estimate Equation (2) using a Tobit regression, where we replace *RPE* with *RPE_prop*, and we obtain qualitatively similar results.

VI. SELECTION OF RPE PEERS

The overall benefits of an RPE contract depend critically on how efficiently the RPE peer group is selected.[18] In this section, we first discuss determinants of RPE peer selection. Next, we compare characteristics between firms that are chosen to be in the RPE peer group and those that are not chosen. We then conduct a multivariate regression analysis of the peer-selection process.

Determinants of RPE Peer Selection

Common Risk Exposure

We measure potential peers' common risk exposure based on the similarity in operational strategies and industry environment, proxied by same industry membership (*Same_SIC2* and *Same_SIC3*), the similarity in the firms' scale and thus flexibility in responding to external shocks, proxied by S&P 1500 membership (*SP1500*) and S&P sub-index membership (*Same_SP*), and performance comovement measured by the correlation in stock returns between a potential peer and the RPE firm of interest (*CORR(PeerReturn*, *RPEReturn)*).

Similar Ability

To capture similarity in ability, we construct |*PeerMVE_RPEMVE*| and |*PeerReturn_RPEReturn*|, defined as the absolute differences in market value of equity and annual stock returns, respectively, between a potential peer and the RPE firm of interest.

Self-Serving Bias

Since firms usually select peer groups used in compensation contracts at the beginning of the fiscal year (e.g., Cadman and Carter 2010; Faulkender and Yang 2010), we measure self-serving bias using potential peers' *expected* performance at the beginning of 2006. Specifically, we construct *PeerReturn_IndReturn*, defined as a potential peer's expected stock performance minus the median expected stock performance for the RPE firm's two-digit SIC industry (excluding the RPE firm of interest). We measure expected stock performance as analysts' first consensus 12-month target price forecast (issued within 90 days after the year beginning) minus the prevailing stock price, scaled by the prevailing stock price.

Symbolism

To test for symbolism considerations in the peer-selection process, we construct *PeerSales_IndSales* and *PeerMVE_IndMVE*, defined as the differences in sales and market value of equity, respectively, between a potential peer and the median levels for the RPE firm's two-digit SIC industry (excluding the RPE firm of interest).

Univariate Results

To get a preliminary assessment of which firms are chosen to be members of a RPE peer group, for each RPE firm that discloses self-selected peers, we identify all domestic firms listed on Compustat that are not selected as RPE peers.[19] Recall that about 74 percent of RPE firms use stock returns as their performance metric when implementing RPE contracts (see Table 1, Panel C). These firms are likely to use stock returns as the basis for selecting peer firms (in addition to

[18] We do not examine the peer-selection process of RPE firms that choose published market or industry indices as RPE peer groups (see Table 1, Panel E) because it is not feasible to test the selection of published indices due to difficulty in gathering the performance and composition information for a population of indices. We also do not examine firms' selection between published indices versus self-selected peer groups due to lack of related economic theories.

[19] Some of our sample firms select foreign firms in forming their RPE peer groups. We do not include foreign firms listed on Compustat, because Compustat only includes foreign firms cross-listed in the United States and thus does not provide an unbiased sample for foreign firms. In addition, Compustat collects firms' reported accounting numbers, while cross-listed firms may report financial statements using non-U.S. accounting standards.

other considerations). To more accurately capture firms' peer-selection process, we further limit this analysis to the 135 RPE firms that employ self-selected peer groups and use stock returns as the RPE-based performance metric.[20]

Table 5 provides descriptive statistics on selected peers and unselected peers. As shown in Panel A, selected peers are more likely to operate in the same industry as the RPE firms, and are more likely to belong to the S&P 1500 index and the same S&P sub-index as the RPE firms. Furthermore, performance comovement between potential peers and the RPE firms is much higher for selected peers than for unselected peers. These results suggest that selected peers share common shocks with the RPE firms to a greater extent than do unselected peers. Consistent with firms choosing peers with similar ability, we find that selected peers have closer size and stock performance to RPE firms than unselected peers. Panel A also reveals a potential peer-selection bias, in that selected peers have significantly lower expected stock performance than unselected peers. Finally, consistent with symbolism in setting CEO pay, we find that selected peers have larger sales and larger market capitalization than unselected peers.

Table 5, Panel B compares selected and unselected peers conditional on whether peers are in the RPE firm's industry. One intriguing finding is that *unselected* peers from RPE firms' industry (within-industry peers) display higher performance comovement than *selected* peers outside RPE firms' industry (outside-industry peers). Hence, cross-industry peer selection does not seem to improve performance comovement. Nevertheless, selected outside-industry peers have larger sales and market capitalizations than selected within-industry peers, consistent with RPE firms selecting larger firms outside their own industries for symbolism considerations.

Multivariate Regression Results

To further examine the selection of RPE peers, we estimate the following logistic regression for the 135 RPE firms that disclose self-selected peer groups and use stock returns as the RPE-based performance metric, with standard errors clustered by RPE firm:

$$\begin{aligned} Prob(RPE\_Peer_{ijt}=1) = \Phi(&\alpha_0 + \alpha_1 Same\_SIC2_{ijt-1} + \alpha_2 Same\_SIC3_{ijt-1} + \alpha_3 SP1500_{ijt-1} \\ &+ \alpha_4 Same\_SP_{ijt-1} + \alpha_5 CORR(PeerReturn, RPEReturn)_{ijt-1} \\ &+ \alpha_6 |PeerMVE\_RPEMVE|_{ijt-1} + \alpha_7 |PeerReturn\_RPEReturn|_{ijt-1} \\ &+ \alpha_8 PeerReturn\_IndReturn_{ijt-1} + \alpha_9 PeerSales\_IndSales_{ijt-1} \\ &+ \alpha_{10} PeerMVE\_IndMVE_{ijt-1} + \varepsilon_{ijt}). \end{aligned} \tag{3}$$

In Equation (3), the dependent variable, RPE_Peer_{ijt}, is an indicator variable that equals 1 if a potential peer firm j (i.e., a domestic firm on Compustat) is chosen to be a member of the RPE peer group by RPE firm i for year t, and 0 otherwise.

Given that the relative proportion of selected versus unselected RPE peers is highly unbalanced, the estimated coefficients from logistic regressions could be significantly biased and inefficient (Owen 2007). To alleviate this concern, we follow the commonly used undersampling method to randomly remove unselected peers when estimating logistic regressions (e.g., Chawla et al. 2003). Specifically, for each RPE firm, we randomly select a set of peers from

[20] The untabulated results based on all RPE firms with self-selected RPE peers are qualitatively similar to those reported in Tables 5 and 6. We do not conduct similar tests on RPE firms using accounting performance metrics because the heterogeneity in RPE-based accounting performance metrics does not allow a sufficiently large sample size to examine any individual accounting performance metric. Nevertheless, if we use earnings as a general proxy for RPE-based accounting performance metrics, we obtain qualitatively similar inferences regarding the RPE peer selection for firms using accounting performance metrics in RPE contracts.

TABLE 5

Mean (Median) of Variables Used to Explain RPE Peer Selection for Selected RPE Peers and Unselected RPE Peers

Panel A: Selected versus Unselected RPE Peers

	Selected RPE Peers (n = 1,668)		Unselected RPE Peers (n = 363,618)		Unselected RPE Peers Used in Table 6 (n = 1,668)	
	Mean	**Median**	**Mean**	**Median**	**Mean**	**Median**
Common Risk Exposure						
Same_SIC2	0.634	1	0.034	0	0.032	0
Same_SIC3	0.454	0	0.016	0	0.016	0
SP1500	0.862	1	0.481	0	0.469	0
Same_SP	0.528	1	0.157	0	0.146	0
CORR(PeerReturn, RPEReturn)	0.364	0.474	0.197	0.265	0.186	0.250
Similar Ability						
|PeerMVE_RPEMVE|	12,081	3,222	15,326	4,005	13,738	3,939
|PeerReturn_RPEReturn|	0.244	0.175	0.389	0.294	0.366	0.274
Self-Serving Bias						
PeerReturn_IndReturn	−0.028	−0.029	0.057	−0.002	0.050	−0.002
Symbolism						
PeerSales_IndSales	8,533	2,911	2,303	152	2,223	125
PeerMVE_IndMVE	14,927	3,884	3,605	377	3,402	318

Panel B: Selected versus Unselected RPE Peers Conditional on RPE Firms' Industry

	Within RPE Firms' Industry				Out of RPE Firms' Industry			
	Selected RPE Peers (n = 1,058)		Unselected RPE Peers (n = 12,238)		Selected RPE Peers (n = 610)		Unselected RPE Peers (n = 351,380)	
	Mean	**Median**	**Mean**	**Median**	**Mean**	**Median**	**Mean**	**Median**
Common Risk Exposure								
Same_SIC2	1	1	1	1	0	0	0	0
Same_SIC3	0.716	1	0.490	0	0	0	0	0

(continued on next page)

Panel B: Selected versus Unselected RPE Peers Conditional on RPE Firms' Industry

	Within RPE Firms' Industry				Out of RPE Firms' Industry			
	Selected RPE Peers (n = 1,058)		Unselected RPE Peers (n = 12,238)		Selected RPE Peers (n = 610)		Unselected RPE Peers (n = 351,380)	
	Mean	**Median**	**Mean**	**Median**	**Mean**	**Median**	**Mean**	**Median**
SP1500	0.847	1	0.439	0	0.887	1	0.482	0
Same_SP	0.482	0	0.130	0	0.608	1	0.158	0
CORR(PeerReturn, RPEReturn)	0.412	0.550	0.346	0.452	0.280	0.314	0.192	0.258
Similar Ability								
|*PeerMVE_RPEMVE*|	8,651.800	2,537.870	17,367.370	5,204.560	16,193.140	4,863.120	15,255.730	3,970.910
|*PeerReturn_RPEReturn*|	0.217	0.153	0.301	0.204	0.294	0.231	0.392	0.298
Self-Serving Bias								
PeerReturn_IndReturn	−0.029	−0.029	0.047	−0.005	−0.027	−0.029	0.057	−0.002
Symbolism								
PeerSales_IndSales	5,834.930	1,986.020	1,925.470	146.780	10,247.360	4,984.220	1,918.150	101.423
PeerMVE_IndMVE	9,296.990	2,842.630	3,074.170	396.034	16,336.840	6,141.290	2,619.980	234.333

Selected RPE peers include domestic firms that are chosen by a RPE firm as RPE peers. Unselected RPE peers include all the other domestic firms that are not chosen by a RPE firm as RPE peers. Unselected RPE peers used in Table 6 include a set of randomly generated firms from Unselected RPE peers and are used in Table 6 regressions. We limit to RPE firms that use stock returns as the performance metric and self-selected peers in RPE contracts.

Variables are measured over or at the end of fiscal year 2005 (2006) if a firm's first proxy filed under the SEC's 2006 executive compensation disclosure rules is for fiscal year of 2006 (2007). |*PeerMVE_RPEMVE*| and |*PeerReturn_RPEReturn*| are winsorized at top one percentiles. *PeerReturn_IndReturn*, *PeerSales_IndSales*, and *PeerMVE_IndMVE* are winsorized at top and bottom one percentiles. The mean differences between Selected RPE peers and Unselected RPE peers are all statistically significant at less than 1 percent level based on two-tailed t-tests (untabulated).

Variable Definitions:

Same_SIC2 and *Same_SIC3* = 1 if the RPE firm and the potential peer share the same two-digit and three-digit SIC code, respectively, and 0 otherwise;

SP1500 = 1 if the potential peer is in the S&P 1500 index, and 0 otherwise;

Same_SP = 1 if the potential peer is in the same S&P 1500 sub-index (S&P 500, Mid-Cap 400, and Small-Cap 600) as the RPE firm, and 0 otherwise;

CORR(PeerReturn, RPEReturn) = Pearson correlation of annual stock returns between the RPE firm and the potential peer over the prior five years;

|*PeerMVE_RPEMVE*| and |*PeerReturn_RPEReturn*| = absolute values of the pair-wise differences between market value of equity and annual stock returns for the potential peer and market value of equity and annual stock returns for the RPE firm; and

PeerReturn_IndReturn, *PeerSales_IndSales*, and *PeerMVE_IndMVE* = pair-wise differences between analysts' forecasted annual stock returns, sales, and market value of equity for the potential peer and median analysts' forecasted annual stock returns, sales, and market value of equity, respectively, for the RPE firm's industry (two-digit SIC code) excluding the RPE firm of interest (requires at least ten firms other than the RPE firm in the industry).

TABLE 6

Logistic Regression of Peer Firm Selection for S&P 1500 Firms Using Self-Selected Peers in Relative Performance Evaluation and Stock Returns as the Performance Metric

Panel A: All RPE Peers

Independent Variables	Predicted Sign	Coefficient (p-value)	Change in Probability Q1 versus Q3 Values
Common Risk Exposure			
Same_SIC2	+	2.394 (<0.001)	0.416
Same_SIC3	+	1.005 (<0.001)	0.232
SP1500	+	0.525 (<0.001)	0.128
Same_SP	+	0.940 (<0.001)	0.219
CORR(PeerReturn, RPEReturn)	+	0.246 (0.001)	0.049
Similar Ability			
|*PeerMVE_RPEMVE*|	−	−0.013 (<0.001)	−0.033
|*PeerReturn_RPEReturn*|	−	−0.178 (0.197)	−0.014
Self-Serving Bias			
PeerReturn_IndReturn	−	−0.921 (<0.001)	−0.038
Symbolism			
PeerSales_IndSales	+	0.019 (0.016)	0.024
PeerMVE_IndMVE	+	0.017 (<0.001)	0.028
Intercept		−1.416 (<0.001)	
Number of selected/unselected peers		1,668/1,668	
Percent concordant/discordant		92.4/7.5	
Pseudo R^2		0.668	
Wald χ^2 (p-value)		895.429 (<0.001)	

Panel B: Conditional on RPE Firms' Stock Performance

		Underperforming RPE Firms		Outperforming RPE Firms	
Independent Variables	Predicted Sign	Coefficient (p-value)	Change in Probability Q1 versus Q3 Values	Coefficient (p-value)	Change in Probability Q1 versus Q3 Values
Common Risk Exposure					
Same_SIC2	+	2.841 (<0.001)	0.445	2.068 (<0.001)	0.388

(continued on next page)

Panel B: Conditional on RPE Firms' Stock Performance

		Underperforming RPE Firms		Outperforming RPE Firms	
Independent Variables	**Predicted Sign**	**Coefficient (p-value)**	**Change in Probability Q1 versus Q3 Values**	**Coefficient (p-value)**	**Change in Probability Q1 versus Q3 Values**
Same_SIC3	+	0.633	0.153	1.362	0.296
		(0.018)		(<0.001)	
SP1500	+	0.213	0.053	1.094	0.249
		(0.093)		(<0.001)	
Same_SP	+	1.127	0.255	0.777	0.185
		(<0.001)		(<0.001)	
CORR(PeerReturn, RPEReturn)	+	0.089	0.016	0.267	0.057
		(0.442)		(0.004)	
Similar Ability					
|PeerMVE_RPEMVE|	−	−0.013	−0.037	−0.016	−0.034
		(0.001)		(0.069)	
|PeerReturn_RPEReturn|	−	0.072	0.005	−0.312	−0.026
		(0.724)		(0.132)	
Self-Serving Bias					
PeerReturn_IndReturn	−	−1.337	−0.051	−0.588	−0.026
		(<0.001)		(0.018)	
Symbolism					
PeerSales_IndSales	+	0.020	0.027	0.020	0.023
		(0.163)		(0.012)	
PeerMVE_IndMVE	+	0.014	0.028	0.019	0.029
		(0.026)		(0.008)	
Intercept		−1.303		−2.058	
		(<0.001)		(<0.001)	
Number of selected/ unselected peers		792/792		876/876	
Percent concordant/ discordant		93.1/6.7		92.1/7.8	
Pseudo R^2		0.690		0.665	
Wald χ^2		436.588		446.964	
(p-value)		<0.001		<0.001	

Panel C: RPE Peers Not Used For Compensation Benchmarking Purposes versus Compensation-Benchmarking Peers Not Used for RPE Purpose

	RPE Peers Not Used for Pay Benchmarking		Benchmarking Peers Not Used for RPE	
Independent Variables	**Coefficient (p-value)**	**Change in Probability Q1 versus Q3 Values**	**Coefficient (p-value)**	**Change in Probability Q1 versus Q3 Values**
Common Risk Exposure				
Same_SIC2	2.566	0.429	2.104	0.391
	(<0.001)		(<0.001)	

(continued on next page)

Panel C: RPE Peers Not Used For Compensation Benchmarking Purposes versus Compensation-Benchmarking Peers Not Used for RPE Purpose

	RPE Peers Not Used for Pay Benchmarking		Benchmarking Peers Not Used for RPE	
Independent Variables	**Coefficient (p-value)**	**Change in Probability Q1 versus Q3 Values**	**Coefficient (p-value)**	**Change in Probability Q1 versus Q3 Values**
Same_SIC3	2.044	0.385	0.545	0.133
	(<0.001)		(0.119)	
SP1500	0.443	0.109	0.702	0.169
	(0.043)		(<0.001)	
Same_SP	0.852	0.201	0.804	0.191
	(<0.001)		(<0.001)	
CORR(PeerReturn, RPEReturn)	0.259	0.051	−0.521	−0.103
	(0.068)		(<0.001)	
Similar Ability:				
|PeerMVE_RPEMVE|	−0.021	−0.052	−0.009	−0.028
	(0.011)		(0.002)	
|PeerReturn_RPEReturn|	−0.306	−0.024	0.570	0.045
	(0.193)		(0.004)	
Self-Serving Bias				
PeerReturn_IndReturn	−1.236	−0.049	1.338	0.056
	(<0.001)		(<0.001)	
Symbolism				
PeerSales_IndSales	0.007	0.010	0.003	0.005
	(0.610)		(0.688)	
PeerMVE_IndMVE	0.022	0.036	0.013	0.031
	(0.027)		(<0.001)	
Intercept	−1.349		−1.751	
	(<0.001)		(<0.001)	
Number of selected/unselected peers	511/511		472/472	
Percent concordant/discordant	89.3/10.6		90.4/9.5	
Pseudo R^2	0.578		0.608	
Wald χ^2	228.344		229.045	
(p-value)	<0.001		<0.001	

The sample includes selected peers and unselected peers (generated from random sampling) for the 135 RPE firms choosing self-selected peer groups and using stock returns as the performance metric in relative performance evaluation. The dependent variable is *RPE_Peer*, which is equal to 1 if the firm (one Compustat domestic firm) is chosen as a peer for relative compensation evaluation purposes, and 0 otherwise. See Table 5 for the other variable definitions. "Underperforming (Outperforming) RPE Firms" include RPE firms with below-industry (above-industry) stock returns during the year prior to the SEC's new executive compensation disclosure rules. *|PeerMVE_RPEMVE|* and *|PeerReturn_RPEReturn|* are winsorized at the top one percentiles. *PeerReturn_IndReturn*, *PeerSales_IndSales,* and *PeerMVE_IndMVE* are winsorized at top and bottom one percentiles. For ease of exposition, coefficients on *PeerSales_IndSales, PeerMVE_IndMVE*, and *|PeerMVE_RPEMVE|* are multiplied by 1,000. The "Change in Probability" columns show the change in the probability of being selected as a RPE peer as the result of moving from the first to the third quartile value of the variable of interest (for indicator variables, moving from 0 to 1), holding all other variables constant at their mean values (for indicator variables, the benchmark probability is determined with the 0 value). Two-tailed p-values in parentheses are based on clustered standard errors at the RPE firm level. The reported p-values in parentheses are two-tailed.

domestic firms on Compustat that are not chosen by the RPE firm. For each RPE firm, we require

that the number of unselected peers in the random sample be equal to the number of selected peers. As reported in Table 5, Panel A, unselected peers generated from random sampling display similar characteristics to those of the population of unselected peers. To ensure the robustness of our results, we generate 100 random samples for each regression reported in Table 6, and find consistent results across each random generation.[21]

Table 6, Panel A reports results from estimating Equation (3). Consistent with the efficient-contracting view, firms that belong to the same industry as the RPE firm, firms included in the S&P 1500 index and from the same S&P sub-index as the RPE firm, and firms with higher return comovement are more likely to be chosen as RPE peers. We also find some evidence that similar firms are more likely to be selected as RPE peers—firms with closer size to the RPE firm are more likely to serve as RPE peers. With respect to selection bias, we find statistically significant evidence that firms with lower expected stock performance are more likely to be chosen as RPE peers. In line with the symbolism considerations, firms with larger sales and larger market capitalization than the RPE firm's industry median are more likely to be chosen as RPE peers.

Among all the factors affecting the peer-selection process, industry membership seems to have the greatest economic significance. For instance, firms from the same two-digit SIC industry as the RPE firm are about 42 percent more likely to be selected as RPE peers than firms are from different industries. S&P index membership also has substantial influence over the likelihood of being selected as a RPE peer. Performance comovement, size similarity, self-serving bias, and symbolism variables have smaller, but still significant, economic effects on RPE peer selection. While the peer-selection bias seems to have small economic significance, we are uncertain about the ultimate impact of peer-selection bias on executive compensation. Taken together, our findings support both efficient-contracting and rent-seeking behavior in the RPE peer-selection process.

We next examine the potential influence of RPE firms' performance on the relative importance of efficient contracting versus rent extraction in peer selection. We posit that underperforming managers are more concerned about job security and pay reduction, and, hence, have stronger incentives to manipulate the peer-selection process. To examine this issue, we re-estimate Equation (3) among RPE firms with below industry median versus above industry median stock returns during the year prior to the SEC's new disclosure regulation. The results are reported in Table 6, Panel B. Consistent with our conjecture, we find that the negative relation between potential peers' expected performance and the likelihood of being selected as a RPE peer is more (less) pronounced for RPE firms that underperform (outperform) their industries. Moreover, the positive impact of stock performance comovement over RPE peer selection is significant among outperforming RPE firms but insignificant among underperforming RPE firms. These findings suggest that superior performance leads efficient-contracting considerations to dominate in the RPE peer-selection process, probably due to managers worrying less about job security and pay reduction. In contrast, peer selection by underperforming firms appears to reflect rent-seeking behavior, probably because managers are concerned about reputation loss and pay reduction and thus have stronger incentives to manipulate the peer-selection process.[22]

---

[21] For instance, with respect to results reported in Table 6, Panel A, across the 100 random-sample regressions, the coefficients on *Same_SIC2*, *Same_SIC3*, *SP1500*, *Same_SP*, *|PeerMVE_RPEMVE|*, *PeerReturn_IndReturn*, and *PeerMVE_IndMVE* are statistically significant as predicted for all 100 regressions. The coefficient on *PeerSales_IndSales* is statistically significant as predicted for 95 regressions. The coefficient on *CORR(PeerReturn, RPEReturn)* is statistically significant as predicted for 93 regressions. Finally, the coefficient on *|PeerReturn_RPEReturn|* is statistically significant as predicted for 44 regressions.

[22] Given that selected peers have larger sales and larger market capitalization than unselected peers, our findings may be confounded by size differences between selected and unselected peers. To mitigate this concern, we restrict the peer

RPE Peer Selection versus Compensation-Benchmarking Peer Selection

Many firms use compensation-benchmarking peers in setting the level of executive compensation, for the purpose of gauging an executive's reservation wage in the competitive labor market.[23] Recent studies suggest that firms select highly paid peers to justify excess CEO compensation (Bizjak et al. 2010; Faulkender and Yang 2010). RPE peers and benchmarking peers, however, not only differ in their theoretical justification, but also in the direction of peer-selection bias for peer performance: better performing firms tend to pay higher compensation and thus are more likely to be selected as compensation-benchmarking peers, whereas worse-performing firms are more likely to be selected as RPE peers to boost firms' relative performance and, hence, incentive compensation.

To shed some light on the peer selection for the purpose of RPE versus the purpose of pay benchmarking, we collect compensation-benchmarking peers for the 232 firms that use self-selected RPE peers in 2006. We find that 147 firms use two different self-selected peer groups[24] (one for RPE purposes and the other for pay-benchmarking purposes), with a mean (median) overlapping rate of 72 percent (81 percent), where the overlapping rate is defined as the number of common peers between the two peer groups divided by the size of the RPE peer group.

We then re-estimate Equation (3) by limiting the sample to selected RPE peers *not* included in the compensation-benchmarking peer group, together with randomly sampled peers. Similarly, we re-estimate Equation (3) by limiting the analysis to selected pay-benchmarking peers *not* used for RPE purposes, together with randomly sampled peers. To be consistent with analyses above, we further require firms to use stock returns as the RPE-based performance metric. Table 6, Panel C reports the regression results. For the selection of RPE peers not used for pay benchmarking, we find qualitatively similar results to those reported in Panel A. Interestingly, we observe that performance comovement negatively influences the selection of pay-benchmarking peers. Moreover, peers' expected stock performance has a positive impact on the likelihood of being chosen as a pay-benchmarking peer, consistent with firms favoring better-performing peers to boost executive pay. These results stand in contrast to the RPE peer-selection results, which support the notion that RPE peers and pay-benchmarking peers serve different purposes and their selection reflects different considerations.

Intertemporal Changes in RPE Peer Group Composition

To provide further evidence on RPE peer selection, we analyze intertemporal changes in RPE peer-group composition. If peer selection conforms to efficient contracting, then we expect added peers to exhibit greater ability to remove common risk and greater similarity with the RPE firm

---

population to S&P 1500 firms, and re-estimate Equation (3). This sample restriction ignores the fact that about 30 percent of selected RPE peers do not belong to the S&P 1500 index (see Table 1, Panel F) which potentially lowers the test power. Nevertheless, we find qualitatively similar results as those reported in Table 6, Panel A (untabulated). We also find qualitatively similar results as those reported in Table 6, Panel B. In particular, the coefficient on *PeerReturn-_IndReturn* is significantly negative for underperforming RPE firms, but insignificantly positive for outperforming RPE firms (untabulated).

[23] Consistent with the efficient use of benchmarking peers, recent studies show that benchmarking peers represent a set of companies against which a firm competes for executive talent (Bizjak et al. 2008, 2010; Cadman and Carter 2010; Albuquerque et al. 2009).

[24] Starting with 232 firms, we are able to find 228 firms' pay-benchmarking peer groups from reading CD&A reports. For these 228 firms, we find that 20 firms use either a broader definition of peers (such as an index) or proprietary compensation survey data for pay-benchmarking purposes. For the 208 firms that use self-selected pay-benchmarking peers, the mean (median) size of pay-benchmarking peer groups is 17 (15) peers. Of these 208 firms, 36 firms use the same peer groups for both pay-benchmarking and RPE purposes; 25 firms have RPE peer groups that are a subset of their pay-benchmarking peer groups; 3 firms use completely different peer groups for two purposes; and the remaining 144 firms have some overlap between the two peer groups.

than dropped peers. If, on the other hand, firms choose poorly performing peers when forming RPE peer groups, then we expect added peers to exhibit worse performance than dropped peers.

To identify the sample, we start with the 232 firms that disclose self-selected RPE peers in 2006. We find that 180 firms continue using RPE in 2007, among which 155 firms disclose self-selected RPE peer groups. Of these 155 firms, 22 firms use the same RPE peer groups and 133 firms modify their RPE peer groups in 2007, with 74 percent of 2006 RPE peers retained in 2007. The final sample includes 75 firms that both dropped and added RPE peers in 2007 and have available data on required variables.

Table 7 reports differences in the factors affecting the selection of RPE peers between added and dropped peers. We find that added peers are closer in size to RPE firms than dropped peers, consistent with more (less) similar firms being added to (dropped from) the RPE peer group. Moreover, added peers' industry-adjusted expected performance is significantly negative, whereas dropped peers' industry-adjusted expected performance is insignificant. While this finding supports self-serving bias in RPE peer selection, the contrast between added and dropped peers is statistically insignificant. This weak evidence could reflect low test power due to stable peer-group composition and/or increased public scrutiny of the peer-selection process after the SEC's new regulation prompting firms to restrain opportunistic behavior in peer selection.

VII. CONCLUSION

We examine the explicit use of relative performance evaluation (RPE) and related peer groups based on S&P 1500 firms' first proxy disclosures under the SEC's 2006 executive compensation

TABLE 7

Mean (Median) of Variables Used to Explain RPE Peer Selection for Peers Added, Dropped, and Unchanged in 2007

	Peers Added (1)	Peers Dropped (2)	Peers Unchanged (3)	Paired t-statistic (1)–(2)
Common Risk Exposure				
Same_SIC2	**0.614**	**0.653**	**0.703**	−1.63
Same_SIC3	**0.481**	**0.465**	**0.552**	1.27
SP1500	**0.720**	**0.707**	**0.789**	0.15
Same_SP	**0.590**	**0.507**	**0.643**	1.49
CORR(PeerReturn, RPEReturn)	**0.357**	**0.424**	**0.394**	−1.54
Similar Ability				
|PeerMVE_RPEMVE|	**15,428.350**	**21,772.750**	**15,703.120**	−1.68*
|PeerReturn_RPEReturn|	**0.232**	**0.253**	**0.222**	−0.89
Self-Serving Bias				
PeerReturn_IndReturn	**−0.035**	−0.014	**−0.038**	−1.55
Symbolism				
PeerSales_IndSales	**10,252.390**	**14,182.150**	**11,710.980**	−0.70
PeerMVE_IndMVE	**18,447.670**	**22,862.210**	**19,342.390**	−0.52

* Indicates significance at less than the 10 percent level based on two-tailed paired t-tests of the mean differences.

Bold figures indicate significance level at less than 5 percent.

The sample includes 75 firms that report two different RPE peer groups for 2006 and 2007 and have available data on required variables. All variables are measured at the beginning of 2007. *PeerReturn* is peer firms' expected stock performance (based on analysts' target price forecasts at the beginning of 2007). See Table 5 for the other variable definitions.

disclosure rules. We demonstrate that a lack of knowledge of both RPE peer-group composition and the link between RPE-based performance targets and future peer-performance cloud inferences drawn from implicit tests. These findings highlight the limitations of the implicit approach and underscore the importance of incorporating explicit RPE contract details in testing for RPE use. We also find that firms consider both costs and benefits of RPE as an incentive mechanism when deciding to use RPE. Analyses of self-selected RPE peers lend support to the efficient use of RPE as well as a self-serving bias in forming RPE peer groups.[25] In particular, RPE firms' superior performance leads efficient-contracting considerations to drive the selection of RPE peers, whereas underperforming RPE firms exhibit greater rent-seeking behavior in the RPE peer-selection process.

One limitation of our analyses is that we are unable to identify firms that use RPE implicitly without pre-committing to explicit RPE contracts. Future research could examine both explicit and implicit contracting to better understand RPE as an incentive device. Furthermore, while our study sheds light on target-setting in executive compensation contracts, our analysis likely omits other factors that boards consider in setting performance targets for executives. The peer-selection bias we document, for example, could reflect boards' conscious efforts to balance various factors in setting performance targets (Merchant and Manzoni 1989). Examining interactions among various factors would be an interesting avenue for future research.

APPENDIX A
TIMELINE OF INTRODUCING THE SEC'S NEW EXECUTIVE COMPENSATION DISCLOSURE RULES

October 16, 1992 to January 27, 2006: *Executive Compensation Disclosure*

The 1992 regulation eschewed a mostly narrative disclosure approach adopted in 1983 in favor of formatted tables that capture all executives' compensation.

January 27, 2006: SEC released the proposed rule *Executive Compensation and Related Party Disclosure*

The proposed rule combines a broader-based tabular presentation with improved narrative disclosure supplementing the tables.

January 27, 2006 to April 10, 2006: Open for public comments

SEC had received more than 20,000 comment letters, most of which supported the new disclosure reform.

August 29, 2006: SEC released the final rule *Executive Compensation Disclosure*

August 29, 2006 to October 23, 2006: Open for public comments

Effective Date:

Companies must comply with these disclosure requirements in Forms 8-K for triggering events that occur on or after November 7, 2006 and in Forms 10-K and 10-KSB for fiscal years ending on or after December 15, 2006. Companies other than registered investment companies must comply

[25] Other than RPE peer selection, there exist alternative channels through which firms can influence the achievability of RPE-based performance targets and, hence, executive pay (such as through raising or lowering the hurdle rate of RPE-based performance targets or selecting different RPE-based performance metrics). We acknowledge that our inferences concerning the efficiency or opportunism in the RPE peer selection do not necessarily imply optimal or excessive executive pay, because alternative design features of RPE contracts (such as the hurdle rate or performance metrics) might either introduce greater rent extraction or offset the selection bias in the RPE peer-selection process.

> with these disclosure requirements in Securities Act registration statements and Exchange Act registration statements (including pre-effective and post-effective amendments), and in any proxy or information statements filed on or after December 15, 2006 that are required to include Item 402 and 404 disclosure for fiscal years ending on or after December 15, 2006. Registered investment companies must comply with these disclosure requirements in initial registration statements and post-effective amendments that are annual updates to effective registration statements on Forms N-1A, N-2 (except those filed by business development companies) and N-3, and in any new proxy or information statements, filed with the Commission on or after December 15, 2006. (directly quoted from the SEC)

APPENDIX B
CODING CRITERIA AND EXAMPLES OF PROXY DISCLOSURES ABOUT RELATIVE PERFORMANCE EVALUATION (RPE)

Identifying RPE Firms from Compensation Discussion and Analysis (CD&A) Reports

If a firm states that at least one component of executive compensation (e.g., annual bonus, restricted stock, stock option, long-term incentive payout) is determined based on firm performance relative to a group of peers, then the firm is identified as a RPE firm; otherwise, the firm is identified as a non-RPE firm. To be conservative, we classify firms that comment favorably on the principle of RPE, but do not disclose details on the use of RPE in setting compensation, as non-RPE firms.

Since the use of compensation-benchmarking peer groups is much more prevalent than the use of RPE peer groups in S&P 1500 firms, it is important to distinguish firms using compensation benchmarking from firms using RPE. The SEC defines compensation benchmarking as "benchmarking of total compensation, or any material element of compensation" (SEC Final Rule 33-8732a, Item 402(b)(2)(xiv)). Consistent with this definition, we classify firms using peer groups to benchmark the *level* of executive compensation as pay-benchmarking firms. Consequently, if our reading of the CD&A report for a sample firm suggests that the firm uses peer groups to solely benchmark the *level* of pay without using peer groups to determine the *performance target* when evaluating the firm's performance, then the firm is coded as a non-RPE firm.

Below are excerpts from the CD&A reports in the first proxy statements filed after December 16, 2006 for four companies. The first three companies are classified as RPE firms. The last company is classified as a non-RPE firm due to lack of details about RPE use in compensation contracts. Discussions about RPE use are bolded.

General Electric Co

> *Performance share units (PSUs).* Since 2003, we have compensated the CEO with PSUs in lieu of any other equity incentive compensation… The receipt of shares underlying PSUs is determined entirely by the performance of the company against two key metrics: an internal metric that measures the cash-producing capability of the company and **an external metric that measures the performance of the company against a broad market index**… PSUs will convert into shares of GE stock at the end of the five-year performance period only if the specified performance objectives have been achieved. Half of the PSUs will convert into shares of GE stock only if GE's cash flow from operating activities, adjusted to exclude the effect of unusual events, has grown an average of 10 percent or more per year over the five-year performance period. Otherwise, they will be cancelled. **The remaining PSUs will convert into shares of GE stock only if GE's total shareowner return meets or exceeds that of the S&P 500 over the five-year performance period. Otherwise, they will be cancelled**.

PG&E Corp

In establishing levels of executive compensation, each year the Committee reviews the appropriateness of the comparator groups used to assess the competitiveness of PG&E corporation's compensation programs (Pay Comparator Group) and PG&E Corporation's corporate performance (**Performance Comparator Group**)… The primary comparator group used for purposes of setting 2006 officer compensation consists of all companies listed in the Dow Jones Utility Index and the Standard & Poor's Electrics Index, and all California investor-owned utilities (the "Pay Comparator Group")… This group of companies is broad enough to provide statistical validity and data availability, represents the segment of the market where PG&E Corporation and Pacific Gas and Electric Company recruit officers with industry-specific experience, and is determined on an objective and transparent basis. **For purposes of corporate performance comparisons (including the relative total shareholder return measured for the 2006–2008 performance share award cycle), the Committee uses a subgroup of 12 companies that have similar characteristics and business models as PG&E Corporation (the "Performance Comparator Group"): Ameren Corporation, American Electric Power, CenterPoint Energy, Inc., Consolidated Edison, Entergy Corporation, FPL Group, NiSource Inc., Pinnacle West Capital, Progress Energy, Inc., Southern Company, TECO Energy, and Xcel Energy. This group of companies is a subset of the Pay Comparator Group and, like PG&E Corporation, is focused on core regulated-utility activities with either a distribution or an integrated-utility focus**.

Performance Shares

The payment for performance shares will be in cash and will be calculated by multiplying (1) the number of vested performance shares, (2) the average closing price of PG&E Corporation common stock over the last 30 calendar days of the year preceding the vesting date, and (3) a payout factor based on corporate performance… **There will be no payout for TSR ("total shareholder return") performance below the 25th percentile of the Performance Comparator Group; there will be a 25 percent payout if TSR is at the 25th percentile; there will be a 100 percent payout if TSR is at the 75th percentile; and there will be a 200 percent payout if PG&E Corporation's TSR ranks first in the Performance Comparator Group. If PG&E Corporation's TSR is between the 25th percentile and the 75th percentile, or above the 75th percentile, award payouts will be determined by straight-line interpolation, adjusted to round numbers (i.e., the nearest multiple of five)**.

Cleco Corp

Our Compensation Committee uses two comparator groups to design executive officer compensation plans and track comparable performance of those plans. These groups are referred to as comparator group(s), peer group(s), peers or the competitive market throughout this discussion. The Base Peer Group was selected based on the companies being of approximate size and scope to Cleco (after regression analysis for size differences), employing similar labor and talent pools and having their executive officer compensation data being available to the outside independent consultant who analyzes the market data for the Compensation Committee. The Compensation Committee considers the availability of such detailed market data to be critical in making comparative compensation decisions… **The Incentive Peer Group was selected by the Compensation Committee in order to measure the actual performance results of our incentive plans. The Incentive Peer Group is based on the companies being part of a recognized stock market index, as well as being in the same general industry classification system. This helps ensure the Compensation Committee evaluates our actual incentive plan performance against a group of companies whose scope of operations and market capitalization is similar to Cleco's**. Data from the peer groups are an integral part of the decision process used by the Compensation Committee in determining the design, component parts and levels of awards contained in our executive officer pay programs.

Base Peer Group

For 2006 and 2007, executive officer compensation levels were evaluated using the Base Peer Group. This includes base salary, annual and long-term incentive plan targets, other potential equity awards and total compensation. The Base Peer Group consists of companies that are generally either in the Edison Electric Institute ("EEI") Index or the S&P Small and MidCap Electric Utilities Index. We are included in both indices.

Incentive Peer Group

For 2006, the relative actual performance of the financial measures used in our annual and long-term incentive plans was determined using the Incentive Peer Group. **The Incentive Peer Group consisted of the following 15 companies contained in the S&P Small and MidCap Electric Utilities Index: Allete, Inc.; Central Vermont Public Service; DPL, Inc.; Duquesne Light Holdings, Inc.; El Paso Electric Company; Great Plains Energy Inc.; Green Mountain Power Corporation; Hawaiian Electric Industries; IDACORP, Inc.; Northeast Utilities; Pepco Holdings, Inc.; Sierra Pacific Resources; UIL Holdings Corporation; UniSource Energy Corporation; and Westar Energy, Inc.** The same 15 companies will be used as the Incentive Peer Group in 2007.

Consolidated Graphics Inc

Overall Compensation Philosophy and Policies

Our compensation philosophy regarding members of the Board and the Company's Executive Officers is to maintain compensation policies which align compensation with the Company's overall business strategy, values and management initiatives. The policies are intended to (1) reward individuals for long-term strategic management and enhancement of shareholder value; (2) **support a performance-oriented environment that rewards achievement of internal Company goals and recognizes the Company's performance compared to the performance of similarly situated companies**; (3) attract and retain individuals whose abilities are considered essential to the long-term future and competitiveness of the Company; and (4) align the financial interests of the Company's directors and Executive Officers with those of the shareholders.

REFERENCES

Aggarwal, R., and A. Samwick. 1999. Executive compensation, strategic competition and relative performance evaluation: Theory and evidence. *The Journal of Finance* 54 (6): 1999–2043.

Albuquerque, A. 2009. Peer firms in relative performance evaluation. *Journal of Accounting and Economics* 48 (1): 69–89.

——— 2007. Do growth option firms use less relative performance evaluation? Working paper, Boston University.

———, G. De Franco, and R. Verdi. 2009. Peer choice in CEO compensation. Working paper, Boston University.

Antle, R., and A. Smith. 1986. An empirical investigation of the relative performance evaluation of corporate executives. *Journal of Accounting Research* 24 (1): 1–39.

Bannister, J., and H. Newman. 2003. Analysis of corporate disclosures on relative performance evaluation. *Accounting Horizons* 17 (3): 235–246.

Barro, J., and R. Barro. 1990. Pay, performance, and turnover of bank CEOs. *Journal of Labor Economics* 8 (4): 448–481.

Bebchuk, L., and J. Fried. 2003. Executive compensation as an agency problem. *The Journal of Economic Perspectives* 17 (3): 71–92.

Bertrand, M., and S. Mullainathan. 2001. Are CEOs rewarded for luck? The ones without principals are. *The Quarterly Journal of Economics* 116 (3): 901–932.

Bizjak, J., M. Lemmon, and L. Naveen. 2008. Does the use of peer groups contribute to higher pay and less

efficient compensation? *Journal of Financial Economics* 90 (2): 152–168.
———, ———, and T. Nguyen. 2010. Are all CEOs above average? An empirical analysis of compensation peer groups and pay design. *Journal of Financial Economics* (forthcoming).
Bol, J., T. Keune, E. M. Matsumura, and J. Y. Shin. 2010. Supervisor discretion in target setting: An empirical investigation. *The Accounting Review* 85 (6) : 1861–1886
Bushman, R., Q. Chen, E. Engel, and A. Smith. 2004. Financial accounting information, organizational complexity and corporate governance systems. *Journal of Accounting and Economics* 37 (2): 167–201.
Byrd, J., M. Johnson, and S. Porter. 1998. Discretion in financial reporting: The voluntary disclosure of compensation peer groups in proxy statement performance graphs. *Contemporary Accounting Research* 15 (1): 25–52.
Cadman, B., and M. Carter. 2010. Compensation peer groups and their relation with CEO compensation. Working paper, The University of Utah.
———, ———, and S. Hillegeist. 2010. The incentives of compensation consultants and CEO pay. *Journal of Accounting and Economics* 49 (3): 263–280.
Carter, M., C. Ittner, and S. Zechman. 2010. Explicit relative performance evaluation in performance-vested equity grants. *Review of Accounting Studies* 14 (2–3): 269–306.
Chawla, N. V., N. Japkowicz, and A. Kolcz. 2003. *Learning from Imbalanced Data Sets*. Proceedings of the ICML 2003 Workshop, Ottawa, Canada.
Cremers, K., and V. Nair. 2005. Governance mechanism and equity prices. *The Journal of Finance* 60 (6): 2859–2894.
DeFond, M., and C. Park. 1999. The effect of competition on CEO turnover. *Journal of Accounting and Economics* 27 (1): 35–56.
Dikolli, S., C. Hofmann, and T. Pfeiffer. 2010. Peer performance aggregation and relative performance evaluation. Working paper, Duke University.
Dye, R. 1984. The trouble with tournaments. *Economic Inquiry* 22 (1): 147–149.
Faulkender, M., and J. Yang. 2010. Inside the black box: The role and composition of compensation peer groups. *Journal of Financial Economics* 96 (2): 257–270.
Ferri, F. 2009. Discussion of: Explicit relative performance evaluation in performance-vested equity grants. *Review of Accounting Studies* 14 (2–3): 307–313.
Garvey, G., and T. Milbourn. 2003. Incentive compensation when executives can hedge the market: Evidence of relative performance evaluation in the cross section. *The Journal of Finance* 58 (4): 1557–1582.
———, and T. Milbourn. 2006. Asymmetric benchmarking in compensation: Executives are rewarded for good luck but not penalized for bad. *Journal of Financial Economics* 82 (1): 197–226.
Gerakos, J., C. Ittner, and D. Larcker. 2007. The structure of performance-vested stock option grants. In *Essays on Accounting Theory in Honor of Joel S. Demski*, edited by R. Antle , P. Liang, and F. Gjesdahl. New York, NY: Springer.
Gibbons, R., and K. Murphy. 1990. Relative performance evaluation for chief executive officers. *Industrial & Labor Relations Review* 43 (3): 30–51.
Green, J., and N. Stokey. 1983. A comparison of tournaments and contracts. *The Journal of Political Economy* 91 (3): 349–364.
Hannan, L., R. Krishnan, and D. Newman. 2008. The effects of disseminating relative performance feedback in tournament and individual performance compensation plans. *The Accounting Review* 83 (4): 893–913.
Himmelberg, C., and R. Hubbard. 2000. Incentive pay and market for CEOs: An analysis of pay-for-performance sensitivity. Working paper, Columbia University.
Holmstrom, B. 1982. Moral hazard in teams. *The Bell Journal of Economics* 13 (2): 324–340.
Hvide, H. 2002. Tournament rewards and risk taking. *Journal of Labor Economics* 20 (4): 877–898.
Janakiraman, S., R. Lambert, and D. Larcker. 1992. An empirical investigation of the relative performance evaluation hypothesis. *Journal of Accounting Research* 30 (1): 53–69.
Joh, S. 1999. Strategic managerial incentive compensation in Japan: Relative performance evaluation and product market collusion. *The Review of Economics and Statistics* 81 (2): 303–313.
Knoeber, C., and W. Thurman. 1994. Testing the theory of tournaments: An empirical analysis of broiler

production. *Journal of Labor Economics* 12 (2): 155–179.

Lazear, E., and S. Rosen. 1981. Rank-order tournaments as optimum labor contracts. *The Journal of Political Economy* 89 (5): 841–864.

Leone, A., and S. Rock. 2002. Empirical tests of budget ratcheting and its effect on managers' discretionary accrual choices. *Journal of Accounting and Economics* 33 (1): 43–67.

Matsumura, E. M., and J. Y. Shin. 2006. An empirical analysis of an incentive plan with relative performance measures: Evidence from a postal service. *The Accounting Review* 81 (3): 533–566.

Merchant, K. A., and J.-F. Manzoni. 1989. The achievability of budget targets in profit centers: A field study. *The Accounting Review* 64 (3): 539–558.

Murphy, K. 1999. Executive compensation. In *Handbook of Labor Economics*, Vol. 3, edited by Ashenfleter, O., and D. Card. Amsterdam, The Netherlands: North-Holland.

——— 2000. Performance standards in incentive contracts. *Journal of Accounting and Economics* 30 (3): 245–278.

———, and T. Sandino. 2010. Executive pay and "independent" compensation consultants. *Journal of Accounting and Economics* 49 (3): 247–262.

Nalebuff, B., and J. Stiglitz. 1983. Prizes and incentives: Towards a general theory of compensation and competition. *The Bell Journal of Economics* 14 (1): 21–43.

O'Keeffe, M., W. Viscusi, and R. Zeckhauser. 1984. Economic contests: Comparative reward schemes. *Journal of Labor Economics* 2 (1): 27–56.

Owen, A. 2007. Infinitely imbalanced logistic regression. *Journal of Machine Learning Research* 8 (4): 761–773.

Prendergast, C. 1999. The provision of incentives in firms. *Journal of Economic Literature* 37 (1): 7–63.

Rajgopal, S., T. Shevlin, and V. Zamora. 2006. CEO's outside employment opportunities and the lack of relative performance evaluation in compensation contracts. *The Journal of Finance* 61 (4): 1813–1844.

Rose, N., and A. Shepard. 1997. Firm diversification and CEO compensation: Managerial ability or executive entrenchment? *The Rand Journal of Economics* 28 (3): 489–514.

Rosen, S. 1986. Prizes and incentives in elimination tournaments. *The American Economic Review* 76 (4): 701–715.

Westphal, J., and E. Zajac. 1994. Substance and symbolism in CEO's long-term incentive plans. *Administrative Science Quarterly* 39 (3): 267–390.

Zajac, E., and J. Westphal. 1995. Accounting for the explanations of CEO compensation: substance and symbolism. *Administrative Science Quarterly* 40 (2): 283–308.

THE ACCOUNTING REVIEW
Vol. 86, No. 3
2011
pp. 1045–1085

American Accounting Association
DOI: 10.2308/accr.00000044

Inflation and Nominal Financial Reporting: Implications for Performance and Stock Prices

Yaniv Konchitchki
University of Southern California

ABSTRACT: The monetary unit assumption of financial accounting assumes a stable currency (i.e., constant purchasing power over time). Yet, even during periods of low inflation or deflation, nominal financial statements violate this assumption. I posit that, while the effects of inflation are not recognized in nominal statements, such effects may have economic consequences. I find that unrecognized inflation gains and losses help predict future cash flows as these gains and losses turn into cash flows over time. I also find significant abnormal returns to inflation-based trading strategies, suggesting that stock prices do not fully reflect the implications of the inflation effects for future cash flows. Additional analysis reveals that stock prices act as if investors do not fully distinguish monetary and nonmonetary assets, which is fundamental to determining the effects of inflation. Overall, this study is the first to show that, although inflation effects are not recognized in nominal financial statements, they have significant economic consequences, even during a period in which inflation is relatively low.

Keywords: *inflation; asset pricing; information; financial reporting; abnormal returns; cash flows; capital markets.*

Data Availability: *Data are available from public sources indicated in the text.*

I. INTRODUCTION

The U.S. financial reporting regime is nominal, which assumes no changes to the purchasing power of the dollar over time. That is, the accounting amounts reported in financial statements based on U.S. Generally Accepted Accounting Principles (GAAP) are called nomi-

I am indebted to my dissertation committee members, Mary Barth (chair), Bill Beaver, and Ron Kasznik, for valuable guidance on this study, and to Steven Kachelmeier (senior editor), Mark Trombley (editor), and two anonymous referees for invaluable comments and suggestions. I thank Nissim Aranya, Snehal Banerjee, Scott Joslin, and Gil Sadka for helpful discussions. I also thank Carlos Corona, Mark DeFond, Dan Givoly, Ilan Guttman, Wayne Landsman, Charles Lee, Scott Richardson, Ronnie Sadka, Ron Siegel, Richard Sloan, Mark Soliman, Steve Stubben, K.R. Subramanyam, Dan Taylor, and seminar participants at Barclays Global Investors, Columbia University, Georgetown University, Rice University, Stanford University, University of Miami, University of Minnesota, The University of North Carolina at Chapel Hill, University of Southern California, and the University of Toronto for helpful comments and suggestions. I am thankful for financial support from Stanford University and University of Southern California.

Editor's note: Accepted by Mark Trombley.

Submitted: February 2009
Accepted: September 2010
Published Online: April 2011

nal because they are not adjusted for inflation.[1] Yet, the annual U.S. inflation rate over the past two decades has averaged to 3 percent. When purchasing power is not constant, a nominal reporting system that combines monetary amounts from different periods violates the Monetary Unit accounting assumption of a stable currency (SFAC 5, Financial Accounting Standards Board [FASB] 1984) and decreases comparability both across firms and over time. Although the effects of inflation on financial reporting received considerable attention during the late 1970s when inflation was relatively high, inflationary effects have been mostly ignored in the more modest inflationary environment since that time.

Nominal financial statements, by their nature, do not account for gains and losses attributable to changes in purchasing power over time. For example, whereas the erosion of a firm's monetary assets (e.g., cash) attributable to inflation is a loss to the firm, the erosion of a firm's monetary liabilities (e.g., debt) is a gain. Further, whereas inflation-adjusted amounts of nonmonetary items (e.g., land) accumulate inflationary effects over time to reflect changes in purchasing power, such effects are not recognized in nominal financial statements. The difference between inflation-adjusted and nominal earnings represents unrecognized gains and losses from inflation. In this study, I investigate the economic consequences of omitting these inflation effects from nominal financial statements. Specifically, I test whether unrecognized gains and losses from inflation help predict future cash flows, and whether investors incorporate this information into their investment decisions.

This study is most closely related to inflation accounting studies that investigate the implications of inflation in the U.S. during the 1970s and 1980s (e.g., Beaver 1979; Easman et al. 1979; Beaver et al. 1980; Gheyara and Boatsman 1980; Ro 1980; Watts and Zimmerman 1980; Beaver and Landsman 1983). These studies address questions primarily related to whether the effects of inflation are associated with contemporaneous annual and short-window stock returns. A conclusion from these studies is that inflation-adjusted (or, interchangeably, IA) data are inconsequential for making financial decisions.[2]

I take a different approach by considering the possibility that inflation can have implications over a period longer than the contemporaneous year. In particular, if unrecognized inflation gains and losses (or simply *inflation gains*, for brevity) are realized over time, then they are likely to help predict firms' future cash flows.[3] Moreover, if the stock market does not fully account for such implications for future cash flows, then inflation gains can be associated with future stock returns. Thus, by focusing attention on possible longer-horizon inflation effects, I shed new light on the extent to which the stock market incorporates inflation information, and on the pricing

[1] In the text I refer to inflation because deflation has rarely occurred in the U.S.; however, the study pertains to both inflation and deflation because deflation can be viewed as negative inflation.

[2] This research mainly investigates questions related to the adoption of the Securities and Exchange Commission (SEC)'s Accounting Series Release (ASR) 190 in 1976 and Statement of Financial Accounting Standard (SFAS) 33 in 1979, which are no longer effective, that were mandated in response to high inflation during the period. Another related area follows Modigliani and Cohn (1979), who hypothesize that the stock market is inefficient because it suffers from inflation illusion (e.g., Ritter and Warr 2002; Campbell and Vuolteenaho 2004; Basu et al. 2006; Chordia and Shivakumar 2005). Another area attempts to explain the association between stock returns and unexpected inflation (e.g., Summers 1981; Bernard 1986). Whereas prior studies in this area examine how raw returns vary with respect to unexpected inflation, I investigate how investors process inflation-adjusted information and whether *ex ante* portfolios can be formed to generate future abnormal returns. I also extend prior work by developing a measure that captures the total information content of inflation that is omitted when reporting nominal amounts (e.g., purchasing dates for different asset classes, as explained in the Appendix), and in contrast to prior studies, which analyze the effects of unexpected inflation on realized returns, I analyze the effects of expected inflation on expected returns. Other related areas investigate inflation effects in a variety of contexts, including the bond market, non-U.S. countries with high inflation, employment determination, dividend decision, and inflation tax (e.g., Phelps 1967; Fama and Schwert 1977; Bar-Yosef and Lev 1983; Gordon 2001; Kothari and Shanken 2004; Davis-Friday and Gordon 2005).

[3] I use the terms "inflation gains" and "unrecognized inflation gains" interchangeably throughout the study.

implications of inflation-adjusted data. I also extend prior research by providing insight into how investors process inflation-adjusted information, thereby providing evidence on the mechanism through which inflation impacts future investment decisions.

The first question I address is whether inflation gains can help predict firms' future cash flows. The inflation effects can translate to future cash flows because higher unrecognized inflation gains accumulated in nonmonetary assets should result in higher cash flows from operations when the assets are used (in the case of property, plant, and equipment [PPE]) or sold (in the case of inventory) through many types of business activities, leading to a positive association between unrecognized inflation gains and future cash flows.

I address this question using two tests. In the first test, I investigate how inflation distorts nominal amounts. To do so, I develop an algorithm that extracts inflation-adjusted data from firms' nominal financial statements. More specifically, the algorithm adjusts nominal financial statements for inflation, firm-by-firm, using the distinction between monetary and nonmonetary amounts, the life cycles of assets and liabilities, and the clean surplus relation. An important feature of the algorithm is that it allows inflation effects to vary across firms and over time, capturing the fact that inflation affects firms differently depending on the structure of their assets and liabilities. Further, the algorithm and the inflation-adjusted amounts in this study rely on the historical cost measurement attribute, which has reliability and objectivity advantages. This is the same system of accounting that underlies the current nominal reporting regime, except that amounts are stated in common units. This test evidences considerable cross-sectional and time-series variation in inflation effects. To verify the external validity of the algorithm, I implement it on a sample of Israeli firms for which I hand-collect inflation-adjusted and nominal data—Israeli firms were required to disclose both sets of figures until 2003. This verification analysis (see the Appendix) shows that the algorithm provides reasonable estimates of inflation-adjusted financial statement amounts.

In the second test, I examine whether unrecognized inflation gains are realized over time, and thus can help to predict future performance. In particular, I test whether unrecognized inflation gains improve forecasts of cash flows from operations for horizons from one to four years relative to a benchmark model developed in Barth et al. (2001). I find that unrecognized inflation gains are associated with future cash flows from operating activities in each of the four subsequent years.

Next, I focus on the stock market effects of inflation. If inflation-adjusted information has implications for future cash flows, then investors are likely to incorporate it into their investment decisions. However, using inflation-adjusted data can be costly because inflation-adjusted data are not reported and processing such data is more complicated than processing nominal data (Beaver and Landsman 1983). Because mispricing can arise when information is costly to obtain and process (Grossman and Stiglitz 1980; Ball 1994), even under the semi-strong form of market efficiency, this raises the question of whether investors fully incorporate the implications of inflation information for future cash flows in their equity-pricing decisions. I therefore investigate whether unrecognized inflation gains can be used to generate significant abnormal returns.

Specifically, given my findings that inflation gains have predictive power for future cash flows, future returns are likely to depend on stock market expectations regarding these gains. If the market fully incorporates information about inflation gains, then stock prices will correctly reflect the implications of such effects for future cash flows, leading to no future abnormal returns. Alternatively, if the market does not fully incorporate information about inflation gains, then stocks may be mispriced, leading to possible future abnormal returns. Accordingly, I predict that if investors overestimate (underestimate) the amount of unrecognized inflation gains in the current period, investors will be negatively (positively) surprised when these effects are realized in cash flows, leading to negative (positive) subsequent abnormal returns.

To test this prediction, I first conduct a portfolio-level return test. This test examines returns for portfolios constructed based on inflation-adjusted information, controlling for common risk

factors. More specifically, this test uses portfolios constructed based on inflation gains and investigates the intercepts (i.e., alphas or abnormal returns) obtained from time-series regressions of future returns in excess of the risk-free rate on the related-period Fama-French factor returns. This test reveals significant intercepts for the high and low inflation-based portfolios, with a zero-cost hedge strategy that results in a significant monthly abnormal return of 77–85 basis points, controlling for the Fama-French, momentum, and net operating assets factors. Thus, it appears that stock prices do not fully reflect the implications of inflation gains. I also find a negative association between inflation gains and future abnormal returns. This finding indicates that inflation gains are underestimated (overestimated) when current-period inflation gains are low (high), leading to the significantly positive (negative) abnormal future returns as the inflation gains turn into future cash flows. As a check, I also conduct a firm-level test that analyzes the extent to which an inflation-based trading strategy predicts firms' future annual abnormal returns. The firm-level test generates abnormal returns in a manner consistent with the portfolio-level test.

The findings show that, although investors take inflation gains into account, they do not do so correctly. In additional analysis, I conduct two tests to examine whether market pricing errors can be explained by how the stock market processes inflation information. I find that the documented ability to generate future abnormal returns and the negative association between inflation gains and future abnormal returns is consistent with an inflation adjustment argument. Specifically, I find that the presence of negative or positive future abnormal returns is consistent with investors not distinguishing monetary and nonmonetary assets when adjusting for inflation, leading to a loss of information because inflation affects these two classes of assets differently. This finding suggests that stock prices act as if investors "fixate" on aggregate assets when taking inflation into account, which leads to errors when adjusting for the effects of inflation.

Robustness tests show that there is no pattern in risk characteristics across inflation gains portfolios, and that an inflation-based factor is not a priced risk factor. These findings indicate that the abnormal returns I document are not attributable to an omitted inflation-based risk factor. Instead, these findings are consistent with abnormal returns stemming from inflation information being costly to obtain and process and, hence, are consistent with market efficiency under costly information (Beaver 1981).

To summarize, this study is the first to find that (1) unrecognized inflation gains are informative for predicting future cash flows, (2) inflation-based trading strategies are associated with significant abnormal returns, meaning that stock prices do not fully reflect the implications of inflation gains for future cash flows, and (3) stock prices act as if investors do not fully distinguish monetary and nonmonetary assets, which is necessary to determine the effects of inflation. Together, these results suggest that, contrary to prior work that concludes inflation-adjusted data are of little consequence to financial decisions, unrecognized inflation effects have significant economic implications, even during a period in which inflation is relatively low.[4]

Section II discusses related background and develops the study's main tests on cash flows and stock returns. Section III presents the research design. Section IV discusses data and sample statistics. Section V provides the main results. Section VI investigates how investors process inflation information. Section VII presents robustness tests. Section VIII concludes.

---

[4] Although I find that inflation-adjusted amounts are informative for predicting future cash flows, I do not propose a shift to an inflation-adjusted reporting regime. Mandating an inflation-adjusted reporting regime in the U.S. may impose public- and firm-level costs that do not necessarily outweigh the benefits, especially when inflation is low. A normative evaluation of the trade-off associated with such a regime shift is beyond the scope of my study.

II. BACKGROUND AND MAIN TESTS

Inflation is a macroeconomic phenomenon that captures the decrease in purchasing power of a currency unit over time because of a general increase in the prices of goods and services (Beaver and Landsman 1983). The fundamental accounting assumption of a Monetary Unit assumes a stable currency (i.e., constant purchasing power) as the unit of record (SFAC 5, FASB 1984). Under U.S. GAAP, if the only activities of a firm are, for example, purchasing a parcel of land for $100 fifty years ago and purchasing an additional parcel of land for $100 one year ago, the firm recognizes land at $200 in its financial statements. Assuming that the purchasing power changes over time, this results in a loss of information because the parcels were purchased with the same dollar amount, but at points in time with different purchasing power. Thus, although comparability is one of the qualitative characteristics of accounting information (SFAC 1, FASB 1978), mixing dollars from different periods distorts the Monetary Unit assumption and impairs comparability across firms and over time. This raises the question as to the economic effects of such inflation effects.

Implications of Inflation Information for Future Cash Flows

When a firm engages in a transaction, the opportunity cost of the transaction from the perspective of a representative investor is the fixed number of general consumption units the firm gives up. However, financial statements, which present nominal amounts, do not include information about the date of each transaction, and items are recognized in terms of dollars and not in terms of consumption units. Therefore, inflation creates a wedge between a recognized nominal amount and its cost in terms of consumption units. For a firm, nominal amounts mix dollars from periods with different dollar-consumption ratios, which impairs comparability over time. Across firms, nominal reporting further impairs comparability because there is large variation among firms' transaction dates and amounts when purchasing power is not constant.[5]

To extract information about transactions in terms of consumption units, financial statement items can be separated into two classes: monetary and nonmonetary. Monetary items are directly measured on the basis of a fixed number of dollars required for their settlement. Nonmonetary items represent either a historical cost or a right (obligation) to receive (deliver) services for which purchasing power is not constant. Under U.S. GAAP, financial statements do not take into account gains and losses on monetary items (e.g., debt) attributable to inflation. For example, although the erosion of monetary assets (liabilities) is a loss (gain) to the firm, such losses and gains are not recognized in U.S. GAAP-based financial statements. Financial statements also do not take into account the effects of inflation on nonmonetary items (e.g., PPE). For example, whereas inflation-adjusted amounts change over time because of inflation, the amounts recognized for gross PPE do not reflect such changes. The difference between earnings based on the two measures, $IGL = IAEarnings - NominalEarnings$, captures unrecognized inflation gains and losses. The variation in inflation-adjusted earnings, and hence in *IGL*, is a result of the level of and changes in inflation, and the difference across firms and over time in the structure of monetary and nonmonetary items. Thus, firms with similar nominal outcomes can differ in their inflation-adjusted out-

---

[5] An example inspired from *The Economist*'s "Big Mac index" (http://www.economist.com/markets/bigmac/about.cfm) illustrates how inflation can lead nominal accounting amounts to not reflect a transaction's opportunity cost in terms of consumption units. Suppose $100 spent on a parcel of land 50 years ago could purchase 100 Big Macs 50 years ago, but because of inflation, $100 can only purchase 40 Big Macs today. Because the $100 spent 50 years ago is reported as $100 in the current-period's nominal financial statements, the accounting system equates the original amount spent by the firm to the amount required to purchase 40 Big Macs today. That is, the accounting system effectively informs investors that the firm gave up an amount corresponding to 40 Big Macs to purchase the parcel of land. Yet the firm in fact purchased the land for the equivalent of 100 Big Macs. In inflation-adjusted terms, the $100 spent 50 years ago is economically equivalent to $250 (=100*100/40) in today's dollars.

comes, and *IGL* can vary for a firm over time even when the nominal amounts remain constant.[6]

Inflation gains can turn into future cash flows in several ways, depending on firms' activities. Because the underlying economics of firms' activities are the same regardless of how these activities are reported in financial statements, if the nominal financial statements do not fully capture the effects of inflation today, such inflation effects are likely to be realized in future periods, thereby enabling the inflation effects to help predict future performance. The inflation effects can turn into future cash flows from operations (*CFO*) because higher unrecognized inflation gains accumulated in nonmonetary assets should result in higher future *CFO* when the assets are used (in the case of PPE) or sold (in the case of inventory). Further, because inflation is correlated with changes in specific prices, predicting higher future *CFO* from increases in the general price index is consistent with prior literature that shows increases in specific prices result in higher *CFO* (e.g., Aboody et al. 1999).[7]

There are many types of business activities for which inflation gains can turn into future cash flows. Consider, for example, the case of PPE. Suppose a parcel of land acquired for \$100 by the firm 50 years ago and, based on the purchasing power at that acquisition time, the firm gave up 100 Big Macs (i.e., one Big Mac—consumption unit—costs \$1). Suppose also that at the acquisition time, the firm expects to generate 10 percent inflation-adjusted annual yield by renting the land to another firm, i.e., it expects cash yield that is equivalent to ten Big Macs per year. Under a nominal reporting regime, firms' transactions are not linked to consumption units. Thus, when the purchasing power decreases over time, a particular number of consumption units in the past is equivalent to more cash in the present. Accordingly, in future periods, when the purchasing power of the dollar decreases because of inflation, the rent income is expected to increase, leading to higher *CFO*. Also, suppose a firm decides to use its land parcel for its own benefit rather than renting it to another firm. As the purchasing power of the monetary unit decreases over time, the use of the land parcel, which is nonmonetary, allows the firm to avoid paying an increasing amount to rent the parcel from another firm. This allows the firm to avoid cash payments that are increasing over time. Another example is inventory. Inventory is often sold several months or years after it is purchased, depending on its turnover, which is a function of the firm's operations and business cycle. The inventory accumulates inflationary gains that result in higher *CFO* upon the inventory sale. Thus, overall, I expect that inflation gains on PPE and inventory translate to higher future *CFO*.

If inflation gains, *IGL*, can turn into future *CFO*, then I predict a positive association between *IGL* and future *CFO*. I also expect inflation gains to turn into future cash flows over several years. First, inflation gains can be accumulated in nonmonetary assets (e.g., land, plant, and buildings) that usually have life cycles of several years. For example, unrealized inflation gains accumulated in a real estate property that is rented out can translate to *CFO* when the firm renews the lease or has another lessee beyond the one-year horizon. Second, inventory inflation gains can last longer than one year under LIFO, which many U.S. firms use, and when inventory turnover is longer than one year.

Note that inflation-adjusted accounting amounts are not fair value amounts. In contrast to fair value amounts, management discretion and subjectivity do not play a role when adjusting for

---

[6] Inflation can lead to large differences between nominal and inflation-adjusted earnings even when inflation is low. For example, during a period of 6 percent average annual inflation in Israel, the median absolute difference between the two earnings amounts is 38 percent of nominal earnings. Also, during a period of 3 percent average annual inflation in the U.S., using the inflation-adjusted earnings for the U.S. sample, the equivalent ratio is 29 percent.

[7] For example, untabulated results suggest that monthly inflation rates are highly and significantly correlated with monthly changes in major indices of nonmonetary assets (commodity and housing) over the past 59 years of available data in the Global Insight database, with Spearman and Pearson correlations of 0.7 on average.

inflation. When adjusting for inflation, I rely on the same measurement attribute underlying financial statements (mainly modified historical cost) and, thus, the adjustment procedure is objective. This is the same system of accounting as under the current nominal reporting regime, except that amounts are stated based on units of equal meaning. In particular, I adjust nominal amounts using a general price index to obtain a common monetary unit. The resulting (inflation-adjusted) amounts capture the implications of inflation for a representative investor who is interested in maintaining consumption units and is thus exposed to an average basket of goods and services.[8]

The Extent to which Stock Returns Reflect Inflation Information

Models in asset pricing often assume that an asset's price depends on the comovement of its payoff with marginal utility from consumption over time and across states of nature. Nominal reporting affects investors' assessment of firms' activities in terms of consumption units. If inflation gains, which capture firms' activities in terms of consumption units, are informative for predicting future cash flows, then this raises the question of whether the stock market fully incorporates such information when valuing firms' equity. Investors are likely to use information if it has implications for future cash flows. However, the use of inflation-adjusted data is more complicated than the use of nominal data (e.g., SFAC 5, FASB 1984). Furthermore, because classes of assets and liabilities are affected differently by inflation, and the inflation impact varies over time based on the structure and age of a firm's assets and liabilities, obtaining inflation-adjusted information (which is publicly unavailable) requires acquisition and processing costs that do not necessarily outweigh their benefits. These factors suggest that inflation accounting may reflect market efficiency under costly information (Grossman and Stiglitz 1980; Beaver 1981; Ball 1994).

If unrecognized inflation gains turn into future cash flows, then future returns should depend on how investors process information about these effects. Thus, I expect to find no subsequent abnormal returns if investors fully incorporate inflation information into their investment decisions. Alternatively, if investors do not fully incorporate information about inflation gains, stocks may be mispriced, leading to possible abnormal returns. Specifically, if investors overestimate (underestimate) inflation gains in the current period, then investors will be negatively (positively) surprised in the future when these gains turn into cash flows, leading to negative (positive) subsequent abnormal returns.

III. RESEARCH DESIGN

Inflation Adjustment Algorithm

To obtain accounting amounts that capture inflation effects, I develop an algorithm that converts nominal amounts into inflation-adjusted amounts on a firm-by-firm basis for a broad sample of firms.[9] The algorithm uses a constant dollar approach to measure firms' activities as if purchasing power were constant. Further, it uses the same measurement attribute underlying nominal financial statements, restating the nominal amounts using an objective general price index.

---

[8] Another reason that I adjust for inflation using a general price index, rather than specific indices for particular categories of assets is to be consistent with inflationary GAAP, which uses the general price index. Further, adjusting using specific assets requires not-readily available data regarding the composition of all assets within each asset class and the associated specific index that captures the price increase in the specific asset. Also, adjusting using specific industry indices is likely to introduce significant measurement error because a firm's composition of assets varies widely even for firms within the same industry. In fact, adjusting based on the general price index is likely to bias against finding significant abnormal returns and *CFO* predictability from using inflation gains.

[9] Prior studies usually focus on particular countries or short periods because of data limitations. By using an algorithm to adjust nominal amounts for inflation I am able to analyze inflation effects on a broader sample.

The algorithm is described in the Appendix. In brief, the algorithm extracts inflation-adjusted information from GAAP financial statements in three main steps. First, it reconstructs, in inflation-adjusted terms, each firm's financial statements over the firm's life until the reporting date of interest. Because the distinction between monetary and nonmonetary items is key to distinguishing the effects of inflation on different accounting amounts, the algorithm separates monetary from nonmonetary assets and liabilities. Second, because monetary items are measured in financial statements based on the fixed number of dollars required for their settlement, the algorithm uses the monetary amounts from the financial statements as the inflation-adjusted amounts. In contrast, the algorithm adjusts nonmonetary items by estimating purchase dates and using information on asset life cycles and inventory turnover. Third, the algorithm estimates the wealth generated each period, in inflation-adjusted terms, by differencing two successive sets of assets and liabilities and using the clean surplus relation to incorporate other comprehensive income amounts, stock issues, and other events that can be affected by inflation over the period (e.g., dividends).[10]

Implications of Inflation Information for Future Cash Flows

If inflation-adjusted earnings are higher (lower) than nominal earnings in the current period, the unrecognized inflation gains (losses) should result in increased (decreased) future *CFO*. To test for the ability of inflation gains to predict future *CFO*, I extend the framework developed in Barth et al. (2001) to include inflation gains. Specifically, for horizons of one to four years, i.e., for $\tau = 1$ through 4, I estimate the following equation:

$$CFO_{t+\tau} = \theta + \beta \cdot IGL_t + \gamma_1 \cdot CFO_t + \gamma_2 \cdot \Delta AR_t + \gamma_3 \cdot \Delta INV_t + \gamma_4 \cdot \Delta AP_t + \gamma_5 \cdot DEPN_t + \gamma_6 \cdot OTHER_t + \mu_{t+\tau}, \tag{1}$$

where IGL_t denotes inflation gains and losses for year t; *CFO*, *INV*, *AR*, *INV*, *AP*, *DEPN*, and *OTHER* are cash flows from operations, accounts receivable, inventory, accounts payable, depreciation and amortization, and other accruals, respectively; and Δ denotes annual change.

In Equation (1), the coefficient on inflation gains, β, is the coefficient of interest. I predict a positive association between inflation gains and future *CFO* ($\beta > 0$). I also predict β to be positive for more than one year because assets can have life cycles longer than one year.[11] Following Barth et al. (2001), I predict the signs on γ_1, γ_2, γ_3, γ_5, and γ_6 to be positive and the sign on γ_4 to be negative.

By construction, the sum of *CFO* and disaggregated accruals equals nominal earnings, and thus the equation includes time t earnings. This allows me to test the incremental predictive effect of IGL_t beyond time t nominal earnings. I estimate Equation (1) using a pooled regression that includes observations cross-sectionally and over time, as well as using annual cross-sectional regressions. For the pooled regression, I base test statistics on regression residuals clustered by firm and year (Petersen 2009; Gow et al. 2010). I include industry fixed effects in the annual cross-sectional regressions, and I report the mean coefficients of time-series parameter estimates

[10] Also, the adjustment procedure is consistent with inflationary GAAP that is currently active or was active in the past, e.g., International Financial Reporting Standards (IFRS): International Accounting Standards (IAS) 15 and 29, International Financial Reporting Interpretations Committee (IFRIC) 7; Israeli GAAP: Institute of Certified Public Accountants in Israel (ICPA) Statements 36 and 50; and U.S. GAAP: SFAS 33 and ASR 190.

[11] My motivation stems from whether there is a positive association between current-period inflation gains and future cash flows, thus I focus on the direction of the coefficient on inflation gains, rather than its magnitude. At the extreme there could be a one-to-one relation between inflation gains and future cash flows. Yet, predicting the magnitude of this coefficient is outside the scope of my study, and it depends on an array of factors and parameters that could vary widely depending on the assumptions regarding these parameters and for which data are unavailable (e.g., the life cycles of nonmonetary assets, the relation between inflation rate and the increase in specific asset prices and their benefits, and the composition of nonmonetary assets).

obtained from the cross-sectional regressions. I conduct tests on these cross-sectional estimates using the Fama and MacBeth (1973) t-statistic and two Z-statistics. I primarily rely on the robust double-clustered pooled specification because Fama-MacBeth t-statistic and the Z-statistics can be overstated if cross-sectional and time-series correlations exist in the data.[12]

The Extent to which Stock Returns Reflect Inflation Information

Portfolio-Level Approach

If mispricing exists, then a relation between inflation gains and subsequent returns is likely to be evident in portfolio-level returns. To test this prediction, the design focuses on the intercepts from portfolios constructed based on inflation gains. The estimated intercepts permit testing the ability of inflation information to explain systematic differences in the cross-section of stock returns, controlling for common risk factors (Fama and French 1993). In particular, I test whether the intercept for the low *IGL* portfolio is significantly different from the intercept for the high portfolio. To implement the test, I construct ten portfolios such that each period all firm-year observations with low (high) *IGL* are sorted into portfolio one (ten). I then calculate future monthly returns for each portfolio and estimate the following time-series equation at the portfolio level to obtain the portfolio intercepts:

$$R_{p,m} - R_{f,m} = \alpha_p + \beta_{p,MKTRF} \cdot MKTRF_m + \beta_{p,SMB} \cdot SMB_m + \beta_{p,HML} \cdot HML_m + \varepsilon_{p,m}, \tag{2}$$

where $R_{p,m}$ is the portfolio return for portfolio p in month m; $R_{f,m}$ is the one-month Treasury bill rate; and $MKTRF_m$, SMB_m, and HML_m, are the Fama-French factor returns, where *MKTRF* is the excess return on the market, and *SMB* and *HML* are constructed based on *MVE* and *BTM*, respectively. Because inflation gains are estimated annually, I align firms' *IGL* on monthly returns accumulated over the 12 months beginning three months after the fiscal year-end, to allow for dissemination of annual reports information and the associated annual inflation rate.

To test for hedge abnormal return from using an inflation-based trading strategy, I construct a zero-cost investment portfolio that longs the lowest portfolio (portfolio one) and shorts the highest portfolio (portfolio ten). I then regress this zero-cost portfolio's returns on the related-period factor returns. The intercept from this zero-cost hedge regression can be interpreted as a monthly abnormal return on a zero inflation-based hedge strategy that buys portfolio one and sells short portfolio ten.

I also sequentially add as controls the Carhart (1997) momentum factor (*UMD*) and a net operating assets factor (*FNOA*) following Hirshleifer et al. (2004). Hirshleifer et al. (2004) show that the ratio of net operating assets to lagged total assets, which they refer to as balance sheet bloat, is associated with future returns. I add this effect in all return tests to control for the possibility that a relation between Net Operating Assets (*NOA*) and inflation gains can affect the association between inflation gains and future returns. To do this, I first obtain *NOA* following Hirshleifer et al. (2004), and then form a *NOA*-based factor following the procedure described in

[12] Two points regarding this test. First, I focus on the inflation effects on future cash flows because cash flows receive considerable attention by the investment community and are of major importance in valuation of firms (e.g., Hackel et al. 2000; Barth et al. 2001). Cash flows are also more difficult to manipulate compared to earnings. However, inflation can also affect nominal earnings because inflation gains can translate into future nominal earnings in several ways (e.g., inventory holding gains). Second, with respect to the Z-statistics, both Z-statistics control for cross-sectional correlations but Z2 also partially corrects for potential upward bias in Z1 arising from lack of independence of parameters across the regression groups (Barth 1994). $Z1 = [(1/\sqrt{N}) \cdot \Sigma_{j=1}^{N} t_j] / \sqrt{[k_j/(k_j-2)]}$, where t_j is the t-statistic for cross-sectional regression j, k_j is the degrees of freedom, and N is the number of cross-sectional groups; and $Z2 = \text{mean}(t) / [\text{stddev}(t) / \sqrt{(N-1)}]$, where mean($t$) and stddev($t$) respectively refer to the mean and standard deviation across the group estimates.

Fama and French (1993) in forming the *HML* and *SMB* factors. The *FNOA* factor is a factor-mimicking *NOA* portfolio.[13]

To the extent that inflation gains result in higher future cash flows, if investors correctly estimate inflation gains, then I expect the future abnormal return, α_p, to be insignificantly different from zero. Alternatively, if investors underestimate (overestimate) inflation gains, then I expect α_p to be significantly positive (negative), indicating positive (negative) subsequent abnormal returns as these inflation gains turn into higher- (lower-) than-expected future cash flows.

As depicted in Figure 1, I adjust for inflation year t amounts using the nominal amounts and inflation time-series rates until year-end t. Thus, inflation gains are known at year-end t and investors have this information before t+1 abnormal returns begin to accumulate. The association between current-period inflation gains and subsequent abnormal returns therefore depends on how the expected and unexpected components of inflation gains are estimated. Analogous to previous research, the source of the surprise that drives future returns, or the unexpected component of inflation gains at time t (and therefore the unexpected cash flows at t+1), is the difference between inflation gains that are estimated correctly versus inflation gains that are estimated without distinguishing monetary and nonmonetary assets.

In other words, similar to the earnings surprise literature, the source of this study's surprise is the difference between the inflation adjustment that I estimate using the algorithm based on inflationary GAAP versus an adjustment that ignores the monetary-nonmonetary distinction.[14] This type of analysis is analogous to that in Sloan (1996), who examines whether investors adequately distinguish the different persistence of the cash flows and accruals components of earnings in predicting future earnings.[15] Also, my use of future abnormal returns to infer the

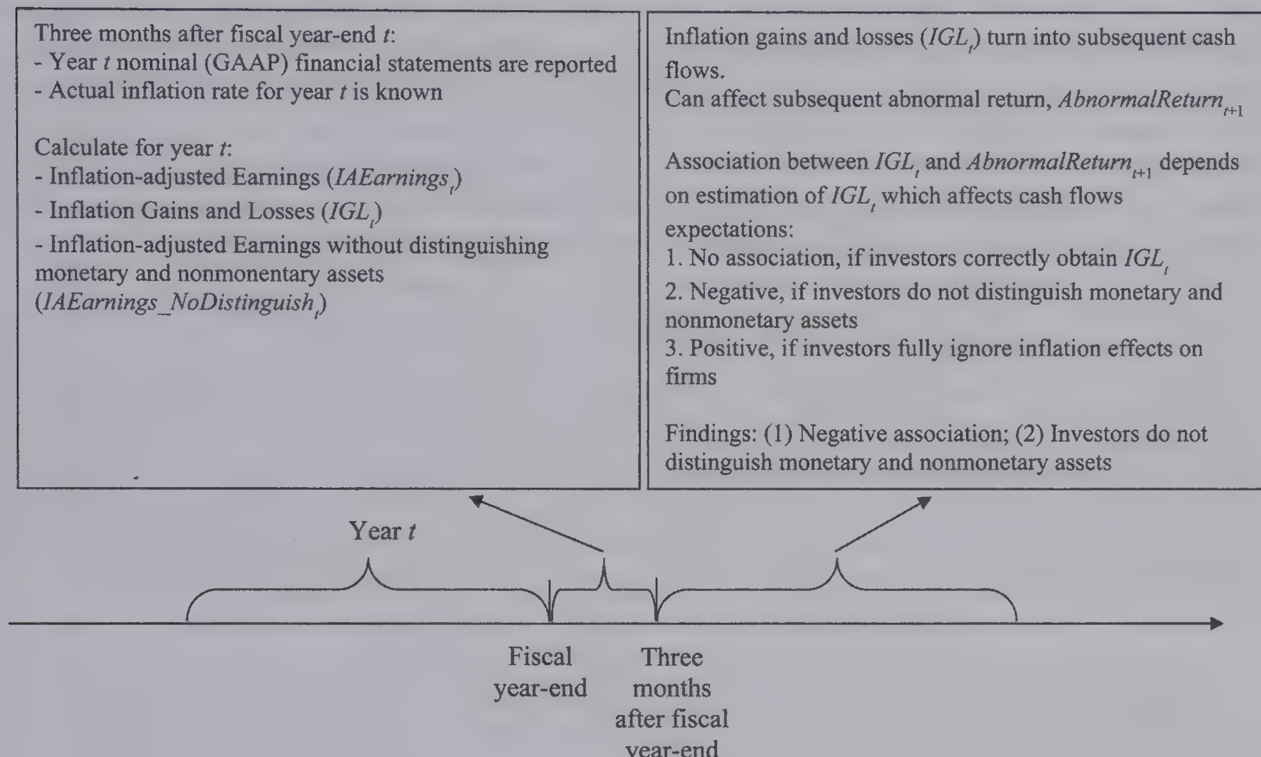
---

[13] Following Hirshleifer et al. (2004), I obtain *NOA* as $NOA = RawNOA \,/\, TotalAssets_{t-1}$, where *RawNOA* = Operating Assets − Operating Liabilities; Operating Assets = Total Assets (Compustat: AT) − Cash and Short-Term Investment (Compustat: CHE); and Operating Liabilities = Total Assets (Compustat: AT) − Debt Included in Current Liabilities (Compustat: DLC) − Long-Term Debt (Compustat: DLTT) − Minority Interests (Compustat: MIB) − Preferred Stocks (Compustat: PSTK) − Common Equity (Compustat: CEQ). Next, I form a *NOA*-based factor following Fama and French (1993). At the end of each month, I sort all observations into two *NOA* groups, where group one (two) includes observations with low (high) *NOA*, and three book-to-market (*BTM*) groups, where group one (three) includes observations with low (high) *BTM*. I then construct six portfolios (L/L, L/M, L/H, H/L, H/M, H/H) from the intersections of the two *NOA* and three *BTM* groups, where the first letter in each of the X/X combinations refers to the *NOA* portfolio (Low, High) and the second letter refers to the *BTM* portfolio (Low, Medium, High). I then calculate monthly value-weighted returns on the six portfolios over the subsequent year, beginning three months after the fiscal year-end. *FNOA* is calculated each month as the average of the monthly returns on the three high *NOA* portfolios (H/L, H/M, and H/H) minus the average of the monthly returns on the three low *NOA* portfolios (L/L, L/M, and L/H).

[14] Therefore, what leads to future abnormal returns is whether investors understand the differential effect of inflation on monetary versus nonmonetary assets, rather than investor understanding of expected versus unexpected inflation. This is different from investors' failure in the current period to distinguish expected and unexpected inflation. Specifically, when I estimate inflation-adjusted amounts, actual inflation is known because it is realized, as shown in Figure 1. Whether actual inflation is fully anticipated or fully unanticipated does not affect my predictions because investors use actual inflation, rather than its expected or unexpected component, to derive inflation gains. Further, the distinction between anticipated and unanticipated inflation is important when examining how current-period earnings changes explain contemporaneous stock price changes, under the notion that stock prices respond to the unanticipated inflation during the year—a setting that was widely used in the research design of inflationary accounting studies during the 1970s-1980s period. In contrast to this contemporaneous setting, my motivation and design are forward-looking, that is, in my study the events flow such that current-period (year t) inflation gains are estimated first, and only in the subsequent period (t+1) do these gains turn into cash flows. The subsequent returns thus do not arise from unanticipated inflation that affects year t inflation gains (see Figure 1).

[15] To the extent that inflation gains are perfectly correlated over time, there may be no surprise component when these inflation effects turn into cash flows over time and thus there may be no theoretical link between inflation gains and future returns. However, when I calculate serial correlations in *IGL* over periods t and t+1, I find Pearson and Spearman correlations of 0.176 and 0.294, respectively. These correlations indicate that inflation gains show only weak persistence over time, as they have low serial correlation. Thus, there is no systematic relation among inflation gains over time. These results are consistent with my expectations, as inflation effects are likely to change over time because there is large variation in the composition of firms' monetary and nonmonetary items over time.

FIGURE 1
Timeline



The figure plots the timeline for analyses throughout the study, and the link between current inflation gains and subsequent abnormal returns. Inflation gains and losses (*IGL*) are estimated after the fiscal year-end, using information from nominal financial statements and actual inflation rates as of the end of the year. *IGL* is defined as the difference between inflation-adjusted earnings and nominal earnings (scaled). Subsequent abnormal returns are accumulated over the year beginning three months after fiscal year-end. Because inflation gains affect the realization of subsequent cash flows, the accuracy of estimating inflation gains affects the accuracy of expectations regarding future cash flows. Subsequent returns are affected depending on actual versus expected inflation gains.

expected versus unexpected components is consistent with prior studies (e.g., Bernard and Thomas 1990; Sloan 1996). For example, similar to Sloan (1996), who infers the expected and unexpected persistence of accruals versus cash flows by examining future abnormal returns, I infer the expected and unexpected components of inflation gains (which lead to a future surprise when these gains affect future cash flows) by examining the patterns in future abnormal returns and inflation gains estimated with and without making the monetary versus nonmonetary distinction.

Firm-Level Approach

As a check on the portfolio-level analysis, I also conduct a firm-level test to discern whether it is possible to earn abnormal returns from inflation-adjusted information. In particular, each year, I form ten portfolios such that firms with the lowest (highest) *IGL* are sorted into portfolio one (ten). Then, for each portfolio-year, I calculate mean abnormal returns, accumulated over the

subsequent year. As before, I align firms' annual amounts to monthly returns in the next 12 months beginning three months after the fiscal year-end.

I examine abnormal returns using four metrics. These are returns adjusted for the value-weighted return on the market (market-adjusted), for the Fama-French factors (Fama-French adjusted), for the Fama-French and momentum factors (Fama-French-*UMD*), and for the Fama-French, momentum, and the net operating assets factors (Fama-French-*UMD*-*FNOA* adjusted). To obtain abnormal returns, I first calculate raw returns by annually compounding each firm's monthly returns. The market-adjusted return is calculated as the annually compounded raw return minus the annually compounded value-weighted return on all NYSE, AMEX, and NASDAQ stocks in CRSP. Next, I estimate the following time-series equation for each firm:

$$R_{i,m} - R_{f,m} = \kappa_i + \beta_{i,MKTRF} MKTRF_m + \beta_{i,SMB} SMB_m + \beta_{i,HML} HML_m + \varsigma_{i,m}. \quad (3)$$

I also estimate an equation similar to Equation (3) after sequentially adding as controls Carhart's (1997) momentum (*UMD*) and the above-defined balance sheet bloat (*FNOA*) factors. Estimation of Equation (3), and its augmented version with *UMD* and *FNOA*, over the sample period yields firm-specific betas, $\beta_{i,MKTRF}$, $\beta_{i,SMB}$, $\beta_{i,HML}$, $\beta_{i,UMD}$, and $\beta_{i,FNOA}$, which I winsorize at the top and bottom one percent. Finally, I obtain abnormal returns by subtracting from raw returns the product of a firm's betas and the respective factor returns, compounded annually.[16,17]

IV. DATA AND SAMPLE STATISTICS

The sample covers U.S. firms with fiscal year-ends over the period 1984 to 2008, a period during which inflation was relatively low (average 3 percent). I obtain nominal accounting variables from the Compustat North America Fundamental Annual database, XPF Format. I obtain monthly raw stock returns from the CRSP Monthly Stock File, and I adjust these for delisting returns following Shumway and Warther (1999) and Beaver et al. (2007).[18] I obtain the risk-free rate, the Fama-French and momentum factors, and portfolio returns from the Fama-French Portfolios and Factors dataset available through the Wharton Research Data Services (WRDS), and the consumer price indices data used in the inflation adjustment procedure from the Global Insight (DRI) dataset's Basic Economics Monthly Series (PRNEW).

To avoid extreme values obtained from deflating by a small denominator or from using a negative book value of equity, I delete observations with total assets, total revenues, or *MVE* lower than \$10 million, and observations with negative book value of equity. To align forecasting tests, and to verify that annual amounts are for a 12-month period, I delete firm-year observations if the firm's fiscal year changed during the year. To mitigate the effects of using penny stocks, I delete stocks with stock price lower than one dollar. The accounting variables are deflated by market

---

16 The firm-level test may lack power due to measurement errors in firm-specific factor betas and *IGL*. The portfolio test, in contrast, allows me to analyze the variation in the cross-section of expected returns. That is, rather than using firm-specific intercepts that depend on unknown firm-level characteristics, the portfolio test conditions on a pre-determined characteristic—inflation gains—and then identifies whether the inflation mispricing effects not explained by the factors vary with this characteristic. Thus, the portfolio test is less subject to this concern.

17 The return analyses involve the interaction of returns, which are nominal, with inflation-adjusted amounts. For comparability with inflation-adjusted amounts, the inflation effects are largely purged from the return metrics both by subtracting the risk-free interest rate from raw returns and by analyzing the incremental effect beyond that contained in the market factor. To check the extent to which the risk-free interest rate is associated with inflation, I compute correlations between the monthly time-series amounts of inflation and risk-free rates. The Spearman and Pearson correlations both are 0.42 and significant, consistent with the risk-free rate absorbing a large portion of the inflation effects from the return metrics. Further, I design the analyses to be on a cross-sectional basis, estimated either monthly or annually. This is because in any given period inflation-adjusted returns are perfectly correlated with nominal returns because inflation is the same for all firms (Beaver and Landsman 1983).

18 Inferences from all return tests are unchanged when I do not adjust for delisting returns.

value of equity (*MVE*) at the beginning of the year, MVE_{t-1}.[19] To mitigate the effects of outliers, I winsorize all variables in Table 1 at the top or bottom one percentile of the deflated value in each year. To reduce measurement error in deriving inflation gains, either from nominal earnings or from using the algorithm, I delete observations in the top or bottom one percentile each year of deflated nominal earnings, inflation-adjusted earnings, and inflation gains. I use the same industry classifications as in Barth et al. (2010). The final sample comprises 64,597 U.S. firm-year observations.

The main explanatory variable of interest in the analysis is *IGL*, or *IAEarnings* minus *NominalEarnings*. *NominalEarnings* is Income before Extraordinary Items as reported in the financial statements, and *IAEarnings* is nominal earnings restated on an inflation-adjusted basis using the algorithm described in the Appendix. To obtain data necessary to estimate Equation (1), I proceed as follows. For observations with fiscal years ending after July 15, 1988, I obtain *CFO* from the statement of cash flows (Barth et al. 1999). Specifically, *CFO* is net cash flow from operations less the accrual portion of extraordinary items and discontinued operations reported on the statement of cash flows, and total operating accruals (*ACCRUALS*) are calculated as *NominalEarnings* minus *CFO*. The following are also calculated based on the statement of cash flows: change in accounts receivable (ΔAR), change in inventory (ΔINV), change in accounts payable and accrued liabilities (ΔAP), and depreciation and amortization (*DEPN*). If ΔAR, ΔINV, or ΔAP is missing, I calculate these items as I do for observations without statements of cash flows. The net of all other accruals, *OTHER*, is $OTHER = NominalEarnings + DEPN - (CFO + \Delta AR + \Delta INV - \Delta AP)$. For observations without a statement of cash flows available, that is, for observations with fiscal year ending before July 15, 1988, I derive *CFO* from accruals using information from the income statement and balance sheet (Dechow et al. 1995; Sloan 1996). To do so, ΔAR, ΔINV, and ΔAP are the change in the applicable balance sheet account (accounts receivable, inventory, and accounts payable plus accrued expenses, respectively), and *DEPN* is depreciation and amortization as reported in the income statement. $ACCRUALS = \Delta AR + \Delta INV - \Delta AP - DEPN + OTHER$, where $OTHER = [(\Delta CA - \Delta CASH) - \Delta CL] - (\Delta AR + \Delta INV - \Delta AP)$; ΔCA, $\Delta CASH$, ΔCL are the respective changes in current assets, cash/cash equivalents, and current liabilities. Then, $CFO = NominalEarnings - ACCRUALS$. Additional variables I use include total assets (*TotalAssets*), net sales (*Revenues*), gross and net PPE (*GrossPPE* and *NetPPE*), and *MVE*, calculated as fiscal year close price multiplied by common shares outstanding. The Appendix includes details on the Israeli data I use for the validation analysis.

Table 1 reports descriptive statistics for select variables. *TotalAssets*, *Revenues*, and *MVE* are presented in million U.S. dollars; *Book-to-Market* and $TotalAssets\,/\,MVE_{t-1}$ are given as ratios; $\beta_{i,MKTRF}$, $\beta_{i,SMB}$, $\beta_{i,HML}$, and $\beta_{i,UMD}$ are the firms' Fama-French and momentum betas, estimated based on Equation (3); and all other variables are deflated by MVE_{t-1}. Panel A reports statistics based on pooled firm-year observations. It shows that the mean (median) MVE_{t-1} is \$2.17 billion (\$217 million). *TotalAssets* are usually larger than MVE_{t-1} with a mean (median) ratio of 2.23 (1.30). The sign on the mean and median variables used in Equation (1) are generally consistent with prior research (e.g., Barth et al. 2001), with a positive sign on *CFO*, *NominalEarnings*, ΔAR, ΔINV, and *DEPN*, and a negative sign on ΔAP, *ACCRUALS*, and *OTHER*. Also, the means of *NonMonAssets* (which consists of *NetPPE*, *INV*, and *Intangibles*) and *NetMonItems* (monetary assets minus monetary liabilities), which are drivers in the inflation adjustment, respectively equal

[19] I use *MVE* as the deflator in the cash flows prediction model for two reasons. First, it is consistent with prior research. Second, my goal is to control for scale differences using a measure that is uncorrelated with my algorithm to avoid spurious results driven from the error in the undeflated models being correlated with the regressors. In the derivation of inflation-adjusted earnings, the algorithm uses several accounting amounts of the firm, so deflating by accounting items (e.g., total assets) may introduce spurious correlations with the accounting variables when predicting cash flows.

TABLE 1
Descriptive Statistics

Panel A: Pooled Firm-Year Observations

	Mean	Median	Std. Dev.	25th Pctl	75th Pctl
TotalAssets	3,874	287	34,430	89	1,153
Revenues	2,012	262	8,790	81	1,000
MVE_{t-1}	2,174	217	12,222	67	874
Book-to-Market	0.61	0.52	0.43	0.32	0.78
$TotalAssets / MVE_{t-1}$	2.23	1.30	4.04	0.71	2.41
CFO	0.12	0.09	0.19	0.04	0.18
ΔAR	0.03	0.01	0.13	0.00	0.04
ΔINV	0.02	0.00	0.09	0.00	0.02
ΔAP	−0.02	−0.01	0.13	−0.03	0.01
DEPN	0.07	0.04	0.09	0.02	0.09
OTHER	−0.08	−0.03	0.27	−0.09	0.01
ACCRUALS	−0.08	−0.04	0.20	−0.12	0.00
NominalEarnings	0.05	0.06	0.13	0.02	0.09
IAEarnings	0.03	0.04	0.17	0.00	0.08
IGL	−0.02	−0.01	0.09	−0.04	0.00
NOA	0.68	0.68	0.34	0.49	0.83
NetPPE	0.58	0.26	0.85	0.09	0.70
INV	0.26	0.10	0.46	0.01	0.30
Intangibles	0.18	0.04	0.36	0.00	0.18
OA	1.05	0.50	1.79	0.27	0.94
TotalLiabilities	1.33	0.60	2.13	0.23	1.45
ReExOCI	0.22	0.24	0.69	0.05	0.48
O	0.09	0.00	0.53	−0.03	0.00
CommonDividends	0.01	0.00	0.02	0.00	0.02
PreferredDividends	0.00	0.00	0.01	0.00	0.00
NewIssues	0.05	0.01	0.15	0.00	0.03
NonMonAssets	1.02	0.62	1.26	0.25	1.31
NetMonItems	−0.22	0.01	0.93	−0.44	0.24

(continued on next page)

Panel A: Pooled Firm-Year Observations

	Mean	Median	Std. Dev.	25th Pctl	75th Pctl
$\beta_{i,MKTRF}$	1.01	0.99	0.49	0.70	1.28
$\beta_{i,SMB}$	0.70	0.62	0.72	0.24	1.05
$\beta_{i,HML}$	0.19	0.33	0.81	−0.17	0.66
$\beta_{i,UMD}$	−0.16	−0.11	0.46	−0.35	0.08

Panel B: Means and Standard Deviations by Industry

	TotalAssets		*NetPPE*		*INV*		*Intangibles*	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Chemicals	2,794	8,319	0.54	0.61	0.27	0.29	0.18	0.28
Computers	1,383	6,560	0.18	0.29	0.12	0.23	0.11	0.24
Durable Manufacturers	1,754	13,449	0.44	0.61	0.34	0.39	0.17	0.31
Extractive Industries	5,222	17,559	1.31	1.05	0.10	0.26	0.05	0.15
Financial Institutions	21,956	112,733	0.13	0.32	0.19	0.58	0.15	0.29
Food	3,699	9,700	0.69	0.86	0.36	0.52	0.26	0.46
Real Estate, Insurance	1,093	3,078	0.49	1.12	0.17	0.50	0.11	0.32
Mining, Construction	1,161	2,369	0.63	0.79	0.70	0.97	0.12	0.28
Other	19,742	90,020	0.54	0.69	0.28	0.46	0.19	0.39
Pharmaceuticals	2,441	8,445	0.16	0.23	0.08	0.15	0.10	0.23
Retail	1,446	5,411	0.59	0.75	0.57	0.68	0.17	0.36
Services	909	2,470	0.66	1.00	0.08	0.23	0.29	0.46
Textile/Print/Publish	1,410	3,380	0.63	0.70	0.35	0.43	0.24	0.45
Transportation	5,784	17,753	1.20	1.27	0.07	0.23	0.39	0.65
Utilities	4,645	7,831	1.81	1.01	0.08	0.12	0.08	0.27

(continued on next page)

	NonMonAssets		*NetMonItems*		*PPE(t)*		*IAEarnings*		*IGL*	
	Mean	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
Chemicals	0.97	0.89	−0.22	0.67	61.95	16.31	0.03	0.16	−0.03	0.10
Computers	0.39	0.57	0.21	0.44	46.02	16.70	−0.01	0.16	−0.01	0.06
Durable Manufacturers	0.93	1.04	−0.11	0.72	60.49	16.50	0.03	0.16	−0.02	0.08
Extractive Industries	1.49	1.32	−0.63	0.95	59.95	13.59	0.02	0.22	−0.02	0.14
Financial Institutions	0.55	1.24	0.31	1.00	50.57	13.51	0.06	0.14	−0.01	0.07
Food	1.31	1.46	−0.47	0.97	64.51	15.42	0.05	0.15	−0.03	0.10
Real Estate, Insurance	0.83	1.66	0.15	1.26	48.17	14.48	0.02	0.21	−0.02	0.11
Mining, Construction	1.56	1.51	−0.59	1.06	54.95	16.13	0.03	0.20	−0.03	0.10
Other	0.99	1.11	−0.13	0.87	55.93	16.82	0.00	0.20	−0.02	0.09
Pharmaceuticals	0.32	0.45	0.12	0.33	56.68	16.96	−0.02	0.14	−0.01	0.05
Retail	1.33	1.40	−0.47	0.99	60.35	15.68	0.04	0.16	−0.02	0.09
Services	1.04	1.38	−0.25	1.01	51.32	15.84	0.03	0.18	−0.01	0.11
Textile/Print/Publish	1.21	1.22	−0.30	0.84	63.12	15.98	0.04	0.17	−0.03	0.10
Transportation	1.71	1.76	−0.79	1.28	58.73	16.09	0.01	0.24	−0.03	0.16
Utilities	1.99	1.16	−1.10	0.86	53.39	6.62	0.04	0.14	−0.04	0.12

The table presents summary statistics of select variables. *TotalAssets* (Compustat: AT), *Revenues* (Compustat: SALE), and *MVE* (market value of equity) are presented in million U.S. dollars; *Book-to-Market* (book value of equity divided by *MVE*), and $TotalAssets\,/\,MVE_{t-1}$ are ratios; $\beta_{i,MKTRF}$, $\beta_{i,SMB}$, $\beta_{i,HML}$, and $\beta_{i,UMD}$ are firm-specific betas, obtained by regressing a firm's time-series monthly excess return on the Fama-French and momentum monthly factors over the sample period; *NOA* is net operating assets deflated by $TotalAssets_{t-1}$, and it is constructed following Hirshleifer et al. (2004); *PPE(t)* is stated in months; and, all other variables are deflated by MVE_{t-1}. *NominalEarnings* is Income before Extraordinary Items as reported in the financial statements (Compustat: IB); *IAEarnings* is nominal earnings restated to an inflation-adjusted basis using the algorithm; *IGL* = *IAEarnings* − *NominalEarnings*, and it captures unrecognized inflation gains and losses. The risk-free rate and the Fama-French and momentum factors are from the Fama-French Portfolios and Factors dataset available through WRDS. $NOA = RawNOA\,/\,TotalAssets_{t-1}$, where *RawNOA* = Operating Assets − Operating Liabilities, and it is obtained following Hirshleifer et al. (2004). Cash flows from operations (*CFO*) is extracted: directly from the statement of cash flows for firms with fiscal year ending after July 15, 1988, or it is derived for firms with fiscal year ending before July 15, 1988. Specifically, for observations with a statement of cash flows available, *CFO* is net cash flows from operating activities (Compustat: OANCF) less the accrual portion of extraordinary items and discontinued operations (Compustat: XIDOC); *ACCRUALS* is total operating accruals, calculated as $ACCRUALS = NominalEarnings - CFO$; ΔAR, ΔINV, ΔAP are the respective annual changes in accounts receivable (Compustat: RECCH), inventory (Compustat: INVCH), and accounts payable and accrued liabilities (Compustat: APALCH); *DEPN* is depreciation and amortization (Compustat: XDP); and *OTHER* is net of all other accruals, calculated as $NominalEarnings + DEPN - (CFO + \Delta AR + \Delta INV - \Delta AP)$. If Compustat items RECCH, INVCH, or APALCH is missing, ΔAR, ΔINV, or ΔAP is the change in the applicable balance sheet account, accounts receivable (Compustat: RECT), inventory (Compustat: INVT), or accounts payable (Compustat: AP) plus accrued expenses (Compustat: XACC), respectively. For observations without a statement of cash flows available, ΔAR, ΔINV, and ΔAP are the change in the applicable balance sheet account, accounts receivable (Compustat: RECT), inventory (Compustat: INVT), and accounts payable (Compustat: AP) plus accrued expenses (Compustat: XACC), respectively; *DEPN* is calculated as for firms with statements of cash flows available; $ACCRUALS = \Delta AR + \Delta INV - \Delta AP - DEPN + OTHER$, where *OTHER* represents accruals not directly captured by ΔAR, ΔINV, or ΔAP, i.e., *OTHER* equals $[(\Delta CA - \Delta CASH) - \Delta CL] - (\Delta AR + \Delta INV - \Delta AP)$, where ΔCA is the change in current assets (Compustat item: ACT), $\Delta CASH$ is change in cash/cash equivalents (Compustat: CHE), and ΔCL is change in current liabilities (Compustat: LCT). Then, $CFO = NominalEarnings - ACCRUALS$. *NonMonAssets* is the sum of *NetPPE*

(continued on next page)

(Compustat: PPENT), *INV* (Compustat: INVT), and *Intangibles* (Compustat: INTAN). Other variables used in developing the algorithm: *OA* = *TotalAssets* − *NonMonAssets*; *TotalLiabilities* (Compustat: LT + MIB); *ReExOCI* is retained earnings excluding other comprehensive income; *O* is other monetary items in stockholders' equity but not in retained earnings; *CommonDividends* (Compustat: DVC); *PreferredDividends* (Compustat: DVP); *NewIssues* is the change over the year in Common Stock (Compustat: CSTK), Preferred Stock (Compustat: PSTK), and Capital Surplus (Compustat: CAPS); and *NetMonItems* is monetary assets minus monetary liabilities. *PPE(t)* is the expected remaining useful life in months for an average *PPE* item. Industries are based on Barth et al. (2010). The sample includes NYSE, AMEX, and NASDAQ U.S. stocks that are in the intersection of CRSP and Compustat, that have data available to calculate inflation gains and losses, and with fiscal year-ends between 1984 and 2008 (n = 25).

1.02 and -0.22, with respective standard deviations of 1.26 and 0.93, indicating a high dispersion in these measures and in *IGL*. As expected given the broad sample, the mean (median) market beta of 1.01 (0.99). The mean (median) of *NominalEarnings* is 0.05 (0.06), which is higher than the mean (median) of *IAEarnings* of 0.03 (0.04). The mean (median) *IGL* is -0.02 (-0.01) with a standard deviation of 0.09.[20] The standard deviations of *IGL* and *IAEarnings* reveal a large variation in these measures.

Table 1, Panel B reports means and standard deviations for each industry by year. With respect to major nonmonetary items, the panel shows that the utilities, extractive, transportation, food, services, mining, and construction industries are the most capital intensive, with *NetPPE* varying between 0.63 and 1.81 in these industries. Panel B also shows that monetary liabilities are larger than monetary assets in most industries, and that *PPE*(t), the expected remaining useful life (in months) for an average *NetPPE* item, is volatile, with the highest age of average assets in the food, textiles, printing, publishing, chemicals, durable, and retail industries (mean *PPE*(t) varies from 60.35 to 64.51 months in these industries). Even within industries, Panel B reveals high variation in monetary and nonmonetary items, which are drivers for inflation adjustments, suggesting high variation in inflation-adjusted earnings.

V. MAIN RESULTS

Implications of Inflation Information for Future Cash Flows

Table 2 reports results from the cash flows analysis. It shows that the coefficients on aggregated accruals are consistent with Barth et al. (2001). It also reveals that β is significantly positive based on the t-statistics and Z-statistics in the pooled and cross-sectional specifications for all four of the forecasting horizons considered.[21] In particular, based on the pooled and cross-sectional specifications, β is significantly different from 0 and equals 0.05 and 0.07 (0.05 and 0.08; 0.05 and 0.11; 0.10 and 0.13), respectively, at the one- (two-, three-, four-) year horizon. This evidence suggests that inflation gains help predict future *CFO*.

The Extent to which Stock Returns Reflect Inflation Information

Table 3 reports results from the portfolio-level return analysis. The results show that five portfolio intercepts are significant, which indicates that forming portfolios based on inflation gains (*IG*) generates significant abnormal returns. The intercept on the lowest portfolio is 0.0049 ($p < 0.001$), whereas the intercept on the highest portfolio is -0.0036 ($p = 0.002$). The intercept of each portfolio can be interpreted as the monthly abnormal return from buying the specific portfolio and selling short a risk-free asset. The table also reveals that, controlling for the Fama-French factors, the monthly zero-cost hedge return for a portfolio constructed on the difference between the intercepts for portfolios one and ten equals 0.00850 and is significant ($p < 0.001$). The table also reveals that this significant abnormal hedge return holds when controlling for momentum (*UMD*) and net operating assets (*FNOA*) factors in addition to the three Fama-French factors, with significant monthly zero-cost hedge returns of 0.00826 ($p < 0.001$) and 0.00767 ($p < 0.001$), respectively. The findings indicate zero-cost abnormal hedge returns of 0.77 percent to 0.85 percent per month for using inflation information, controlling for common risk factors.

[20] As explained in the Appendix, *IGL* is normalized based on a reference point underlying the adjustment procedure. Because I adjust accounting amounts to be stated in terms of constant dollars to maintain the purchasing power at the estimated purchasing date of nonmonetary items, *IGL* is more frequently negative. Alternatively, *IGL* can be adjusted such that it is more frequently positive. The variation across firms and over time is unchanged under the two alternatives, and so are the cross-sectional results throughout the study.

[21] p-values of 5 percent or less are considered statistically significant, and all significance levels are one-tailed where I have predictions and two-tailed otherwise.

TABLE 2

Implications of Inflation Information for Future Cash Flows

	Prediction	θ	*IGL*	*CFO*	ΔAR	ΔINV	ΔAP	*DEPN*	*OTHER*	Adj.R^2
		?	+	+	+	+	−	+	?	
						CFO_{t+1}				
Pooled	Coeff.	0.02	0.05	0.52	0.25	0.14	−0.02	0.68	0.10	0.381
	t-stat	17.14	2.95	40.31	14.86	6.23	−1.03	31.06	7.73	
	p-value	<0.001	0.002	<0.001	<0.001	<0.001	0.152	<0.001	<0.001	
Cross-Sectional	Mean Coeff.	0.01	0.07	0.51	0.29	0.23	0.09	0.66	0.12	0.402
	FM t-stat	0.92	4.37	17.95	11.81	7.01	2.15	17.55	3.75	
	p-value	0.186	<0.001	<0.001	<0.001	<0.001	0.024	<0.001	<0.001	
	Z1	0.92	7.64	102.58	35.60	23.57	7.45	64.42	18.54	
	p-value	0.186	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	
	Z2	0.67	4.46	12.11	15.37	6.90	2.05	20.68	4.98	
	p-value	0.256	<0.001	<0.001	<0.001	<0.001	0.029	<0.001	<0.001	
						CFO_{t+2}				
Pooled	Coeff.	0.03	0.05	0.52	0.25	0.22	−0.06	0.74	0.10	0.307
	t-stat	20.51	2.77	34.00	12.53	8.46	−2.33	26.50	6.68	
	p-value	<0.001	0.003	<0.001	<0.001	<0.001	0.010	<0.001	<0.001	
Cross-Sectional	Mean Coeff.	0.00	0.08	0.51	0.27	0.26	−0.01	0.71	0.12	0.328
	FM t-stat	−0.15	3.06	27.96	10.48	7.74	−0.33	19.53	5.07	
	p-value	0.441	0.004	<0.001	<0.001	<0.001	0.373	<0.001	<0.001	
	Z1	−0.30	6.47	76.70	25.37	21.37	−0.73	53.92	15.52	
	p-value	0.384	<0.001	<0.001	<0.001	<0.001	0.238	<0.001	<0.001	
	Z2	−0.28	2.80	17.25	11.74	7.67	−0.30	21.57	6.73	
	p-value	0.392	0.006	<0.001	<0.001	<0.001	0.384	<0.001	<0.001	

(continued on next page)

TABLE 2 (continued)

		θ	*IGL*	*CFO*	ΔAR	ΔINV	ΔAP	*DEPN*	*OTHER*	
	Prediction	?	+	+	+	+	−	+	?	**Adj.R²**
						CFO_{t+3}				
Pooled	Coeff.	0.04	0.05	0.52	0.29	0.29	−0.07	0.87	0.10	0.266
	t-stat	22.48	2.18	27.17	11.00	9.22	−2.41	25.57	5.08	
	p-value	<0.001	0.015	<0.001	<0.001	<0.001	0.008	<0.001	<0.001	
Cross-Sectional	Mean Coeff.	0.00	0.11	0.51	0.31	0.32	−0.03	0.82	0.12	0.295
	FM t-stat	−0.10	3.89	19.84	9.24	6.28	−0.98	19.97	4.89	
	p-value	0.461	<0.001	<0.001	<0.001	<0.001	0.171	<0.001	<0.001	
	Z1	−0.25	6.45	58.30	23.14	20.93	−1.87	49.28	12.46	
	p-value	0.403	<0.001	<0.001	<0.001	<0.001	0.040	<0.001	<0.001	
	Z2	−0.20	3.13	16.90	9.58	6.25	−1.21	19.43	6.16	
	p-value	0.422	0.003	<0.001	<0.001	<0.001	0.122	<0.001	<0.001	
						CFO_{t+4}				
Pooled	Coeff.	0.06	0.10	0.56	0.34	0.25	−0.09	1.04	0.10	0.239
	t-stat	21.09	3.90	22.46	9.64	6.67	−2.15	23.49	3.71	
	p-value	<0.001	<0.001	<0.001	<0.001	<0.001	0.016	<0.001	<0.001	
Cross-Sectional	Mean Coeff.	0.00	0.13	0.54	0.39	0.31	−0.04	0.93	0.13	0.270
	FM t-stat	0.04	2.69	15.63	9.35	6.87	−0.89	20.76	4.12	
	p-value	0.484	0.008	<0.001	<0.001	<0.001	0.193	<0.001	<0.001	
	Z1	−0.11	6.62	47.49	21.59	15.21	−1.91	43.71	10.49	
	p-value	0.457	<0.001	<0.001	<0.001	<0.001	0.037	<0.001	<0.001	
	Z2	−0.12	2.36	14.74	9.33	6.71	−0.99	19.03	4.36	
	p-value	0.453	0.016	<0.001	<0.001	<0.001	0.168	<0.001	<0.001	

(continued on next page)

TABLE 2 (continued)

The table reports results from estimating pooled and cross-sectional regressions of the following equation:

$$CFO_{t+\tau} = \theta + \beta \cdot IGL_t + \gamma_1 \cdot CFO_t + \gamma_2 \cdot \Delta AR_t + \gamma_3 \cdot \Delta INV_t + \gamma_4 \cdot \Delta AP_t + \gamma_5 \cdot DEPN_t + \gamma_6 \cdot OTHER_t + \mu_{t+\tau},$$

for $\tau = (1, 2, 3, 4)$.

The pooled (cross-sectional) specification is estimated using double-clustering of regression errors by firm and year (annual cross-sectional regressions and including industry indicator variables that are omitted for brevity). *CFO* is cash flows from operating activities. *IGL* = *IAEarnings* − *NominalEarnings*, deflated by the lagged market value of equity, and it captures unrecognized inflation gains and losses. *NominalEarnings* is Income before Extraordinary Items. *IAEarnings* is nominal earnings restated to an inflation-adjusted basis using the algorithm. The control variables are defined in Table 1 and are based on Barth et al. (2001). Industries are based on the 15 industries in Barth et al. (2010). In the cross-sectional specification: Mean refers to the mean coefficient, averaged across the cross-sectional estimates; FM t-stat refers to the Fama and MacBeth (1973) t-statistic; Z1 $= [(1/\sqrt{N}) \cdot \Sigma_{j=1}^{N} t_j] / \sqrt{[k_j/(k_j-2)]}$, where t_j is the t-statistic for cross-sectional regression j, k_j is the degrees of freedom, and N is the number of cross-sectional groups; and Z2 $= \text{mean}(t) / [\text{stddev}(t) / \sqrt{(N-1)}]$, where mean($t$) and stddev($t$) respectively refer to the mean and standard deviation across the group estimates. p-values for sign-predicted coefficients are obtained from one-tailed t-tests. The sample includes all NYSE, AMEX, and NASDAQ U.S. stocks that are in the intersection of CRSP and Compustat, that have data available to calculate inflation gains and losses, and with fiscal year-ends between 1984 and 2008 (n = 25).

TABLE 3

Abnormal Returns from Using Inflation Information: Portfolio Level

	Coefficient				t-statistic (p-value)					
IGL Portfolio	α_p	β_{MKTRF}	β_{SMB}	β_{HML}	α_p	β_{MKTRF}	β_{SMB}	β_{HML}	Adj. R^2	Mean *IGL*
1	0.0049	1.033	0.776	0.242	4.37	40.14	21.62	6.22	0.893	−0.080
					(<0.001)	(<0.001)	(<0.001)	(<0.001)		
2	0.0046	0.986	0.772	0.226	4.78	44.97	25.24	6.83	0.915	−0.035
					(<0.001)	(<0.001)	(<0.001)	(<0.001)		
3	0.0042	0.992	0.723	0.233	4.46	45.56	23.82	7.09	0.914	−0.026
					(<0.001)	(<0.001)	(<0.001)	(<0.001)		
4	0.0041	1.008	0.745	0.191	4.68	50.08	26.53	6.28	0.929	−0.020
					(<0.001)	(<0.001)	(<0.001)	(<0.001)		
5	0.0030	1.004	0.671	0.145	1.68	46.84	22.43	4.47	0.917	−0.016
					(0.093)	(<0.001)	(<0.001)	(<0.001)		
6	0.0017	0.982	0.621	0.217	1.62	39.99	18.11	5.84	0.884	−0.011
					(0.106)	(<0.001)	(<0.001)	(<0.001)		
7	0.0001	0.982	0.507	0.278	0.11	39.48	14.62	7.41	0.871	−0.007
					(0.913)	(<0.001)	(<0.001)	(<0.001)		
8	−0.0004	1.030	0.508	0.255	−0.38	41.97	14.84	6.89	0.884	0.005
					(0.703)	(<0.001)	(<0.001)	(<0.001)		
9	−0.0013	1.055	0.607	0.212	−1.26	43.29	17.86	5.76	0.897	0.012
					(0.207)	(<0.001)	(<0.001)	(<0.001)		
10	−0.0036	1.086	0.716	0.193	−3.10	41.15	19.43	4.84	0.893	0.068
					(0.002)	(<0.001)	(<0.001)	(<0.001)		
Fama-French Factors:										
Zero-cost hedge (1–10)	0.00850				8.56					
Add *UMD*, No *FNOA*:					(<0.001)					
Zero-cost hedge (1–10)	0.00826				8.19					
Add *UMD*, Add *FNOA*:					(<0.001)					
Zero-cost hedge (1–10)	0.00767				7.71					
					(<0.001)					

(continued on next page)

TABLE 3 (continued)

The table reports results from estimating time-series monthly portfolio regressions. First, each firm-year observation accumulates return over the 12 months beginning three months after the fiscal year-end. Second, ten portfolios are constructed each period with the lowest (highest) *IGL* sorted into portfolio one (ten), rebalanced monthly. *IGL* is inflation-adjusted earnings minus nominal earnings (scaled), and it captures unrecognized inflation gains and losses. Third, average portfolio excess returns, $R_{p,m} - R_{f,m}$, are calculated each month m using all observations in that portfolio, and these monthly portfolio excess returns are regressed on the Fama-French factors (*MKTRF*, *SMB*, *HML*), sequentially adding as controls the Carhart (1997) momentum (*UMD*) and the net operating assets (*FNOA*) factors. The intercepts from these regressions are reported as α_p in the table. The magnitude and statistical tests conducted on the difference between the highest and the lowest portfolios are from regressing zero-cost investment hedge portfolio returns, which are obtained after longing the lowest portfolio and shorting the highest portfolio, on the related-period factors. The risk-free rate, $R_{f,m}$, is the one-month Treasury bill rate. Monthly raw stock returns are obtained from the CRSP Monthly Stock File, and are adjusted for delisting returns. The risk-free rate and the Fama-French and momentum factors from the Fama-French Portfolios and Factors dataset available through WRDS. *NOA* is net operating assets deflated by lagged total assets, constructed as described in Hirshleifer et al. (2004). *FNOA* is a factor-mimicking *NOA* portfolio, formed by first sorting each month m all observations into two *NOA* groups and three book-to-market (*BTM*) groups, and constructing six portfolios from the intersections of the two *NOA* and three *BTM* groups. I then calculate monthly value-weighted returns on the six portfolios over the subsequent year, beginning three months after the fiscal year-end. *FNOA* is then calculated each month as the average of the monthly returns on the three high *NOA* portfolios minus the average of the monthly returns on the three low *NOA* portfolios. p-values are obtained from two-tailed t-tests. The sample includes all NYSE, AMEX, and NASDAQ U.S. stocks that are in the intersection of CRSP and Compustat, that have data available to calculate inflation gains and losses, and with fiscal year-ends between 1984 and 2008 (n = 25).

Untabulated results from a trend regression of the intercepts on ordered portfolios indicate that the decline in intercepts across the ten portfolios is significant ($p < 0.001$), suggesting a significantly negative relation between inflation gains and subsequent abnormal returns.

Table 4 reports results from the firm-level return analysis. The table presents the annual mean abnormal returns, accumulated over the year subsequent to portfolio formation and aggregated over the lowest and highest *IGL* portfolios. Similar to the results from the portfolio-level analysis, the firm-level results show that inflation information generates significant abnormal returns, with statistically significant mean hedge abnormal returns over the sample period that vary between 8.8 percent and 9.8 percent per year across the four abnormal return metrics considered. The results also show that, consistent with the results in Table 3, the low portfolio consistently yields higher subsequent returns than the high portfolio, with positive abnormal returns to a hedge strategy in 22 to 23 of the 25 sample years.[22]

Figure 2 plots the four abnormal return metrics accumulated over the year subsequent to portfolio formation, illustrating the dominance of the low portfolio (solid lines) over the high portfolio (dotted lines) on average in about 90 percent of the sample years. This reflects a time-consistent pattern in future abnormal returns.

Overall, the findings of significant future abnormal returns provide consistent evidence that inflation-adjusted information is not fully incorporated by the stock market.

VI. INVESTOR PROCESSING OF INFLATION INFORMATION

The association between current-period inflation gains (IGL_t) and subsequent abnormal returns ($AbnormalReturn_{t+1}$) depends on how investors estimate IGL_t. Figure 1 illustrates that IGL_t is estimated using information before $AbnormalReturn_{t+1}$ begin to accumulate. If investors correctly estimate IGL_t, then there is nothing unexpected with respect to the future realization of inflation gains in cash flows, such that no future abnormal returns are predicted (and hence no association between IGL_t and $AbnormalReturn_{t+1}$ is predicted). Alternatively, if investors completely ignore inflation, rendering as unexpected the entire future realization of inflation gains in cash flows, then it leads to a future positive surprise when inflation gains are realized. Such a surprise when ignoring inflation gains is commensurate with IGL_t because the higher the (ignored) inflation gains are, the more favorable are future cash flows relative to investors' expectations, leading to a predictable positive association between IGL_t and $AbnormalReturn_{t+1}$.

However, the previous section provides evidence of (1) the existence of future abnormal return, and (2) a negative association between IGL_t and $AbnormalReturn_{t+1}$. Thus, investors appear to neither correctly estimate nor completely ignore inflation gains. This raises the question as to whether, in attempting to adjust for inflation, investors make errors in doing so. Accordingly, I examine how inflation information is processed by investors. Prior studies indicate that investors "fixate" on aggregate amounts without distinguishing components of the aggregate amounts. For example, Sloan (1996) provides evidence consistent with investors "fixating" on aggregate earnings, failing to distinguish the different implications of the accrual and cash flow components of earnings for future performance. Similarly, because inflation affects monetary and nonmonetary assets differently, stocks prices will be affected if investors rely on aggregate amounts instead of distinguishing their different components. I therefore investigate how investors process the implications of inflation gains by examining whether "fixating" on aggregate amounts without distin-

[22] The second column in Table 4 provides the annual inflation rate, calculated based on the PRNEW index from the Global Insight (DRI) dataset's Basic Economics Monthly Series. The inflation rate can be obtained from different data providers with slight differences, yet different measures are often highly correlated. For example, annual inflation rates over my sample period calculated from the Global Insight and World Bank datasets are very close to each other, with a high and significant correlation (correlation = 0.82).

TABLE 4

Future Abnormal Returns from Using Current-Period Inflation Information: Firm Level

					Add Factors			Add *UMD*			Add *FNOA*		
		Market Adjusted Return			Fama-French Adjusted Return			Fama-French-*UMD* Adjusted Return			Fama-French-*UMD-FNOA* Adjusted Return		
Year of Portfolio Construction	Annual Inflation Rate	Low	High	Hedge	Low	High	Hedge	Low	High	Hedge	Low	High	Hedge
1984	0.0355	−0.12	−0.17	0.051	−0.09	−0.16	0.077	−0.06	−0.13	0.069	−0.06	−0.09	0.029
1985	0.0362	−0.02	−0.12	0.100	0.03	−0.05	0.080	0.04	−0.04	0.081	0.06	−0.04	0.105
1986	0.0074	0.03	−0.08	0.115	0.05	−0.05	0.104	0.04	−0.06	0.099	0.05	−0.05	0.104
1987	0.0447	0.03	−0.01	0.044	0.01	−0.03	0.041	0.02	−0.02	0.040	0.02	−0.02	0.046
1988	0.0428	−0.08	−0.12	0.040	0.01	−0.02	0.027	0.03	0.01	0.024	0.03	0.00	0.025
1989	0.0469	−0.02	−0.09	0.064	0.07	0.00	0.077	0.09	0.01	0.079	0.10	0.02	0.078
1990	0.0608	0.11	−0.01	0.127	0.04	−0.07	0.112	0.06	−0.06	0.117	0.07	−0.05	0.119
1991	0.0279	0.07	−0.09	0.157	0.00	−0.12	0.118	0.04	−0.09	0.130	0.04	−0.07	0.116
1992	0.0286	0.12	0.02	0.107	0.04	−0.04	0.085	0.06	−0.02	0.079	0.08	−0.01	0.090
1993	0.0257	0.05	−0.09	0.147	0.09	−0.05	0.143	0.09	−0.05	0.144	0.12	−0.03	0.155
1994	0.0271	0.07	−0.11	0.183	0.07	−0.11	0.186	0.09	−0.08	0.167	0.16	−0.06	0.214
1995	0.0251	−0.08	−0.14	0.058	−0.06	−0.12	0.061	−0.04	−0.09	0.051	−0.02	−0.07	0.047
1996	0.0330	−0.06	−0.08	0.020	−0.08	−0.11	0.032	−0.04	−0.06	0.022	−0.01	−0.05	0.040
1997	0.0147	−0.22	−0.37	0.151	−0.04	−0.13	0.093	0.01	−0.08	0.088	0.01	−0.08	0.088
1998	0.0158	0.36	−0.04	0.400	0.25	−0.15	0.402	0.34	−0.07	0.410	0.40	−0.03	0.425
1999	0.0279	0.26	0.16	0.099	0.21	0.00	0.204	0.23	0.02	0.203	0.28	0.04	0.234
2000	0.0338	0.42	0.22	0.204	0.29	0.12	0.175	0.31	0.13	0.176	0.34	0.13	0.208
2001	0.0129	0.08	0.00	0.077	0.07	0.01	0.052	0.09	0.03	0.063	0.13	0.03	0.106
2002	0.0248	0.56	0.30	0.260	0.30	0.09	0.209	0.29	0.09	0.204	0.31	0.10	0.206
2003	0.0186	−0.07	0.00	−0.074	−0.10	−0.01	−0.088	−0.06	0.01	−0.074	−0.06	0.02	−0.078
2004	0.0354	0.09	0.05	0.047	0.00	−0.04	0.043	0.03	−0.01	0.040	0.03	0.00	0.032
2005	0.0347	−0.04	−0.07	0.029	−0.01	−0.06	0.056	−0.01	−0.06	0.053	0.00	−0.06	0.055
2006	0.0242	−0.09	−0.06	−0.030	0.00	0.04	−0.044	0.04	0.07	−0.036	0.04	0.05	−0.013
2007	0.0446	0.04	−0.04	0.077	0.02	0.00	0.019	−0.01	−0.04	0.027	0.00	−0.04	0.039
2008	−0.0067	0.11	0.11	0.000	0.03	0.07	−0.043	−0.03	0.01	−0.047	−0.04	0.01	−0.048

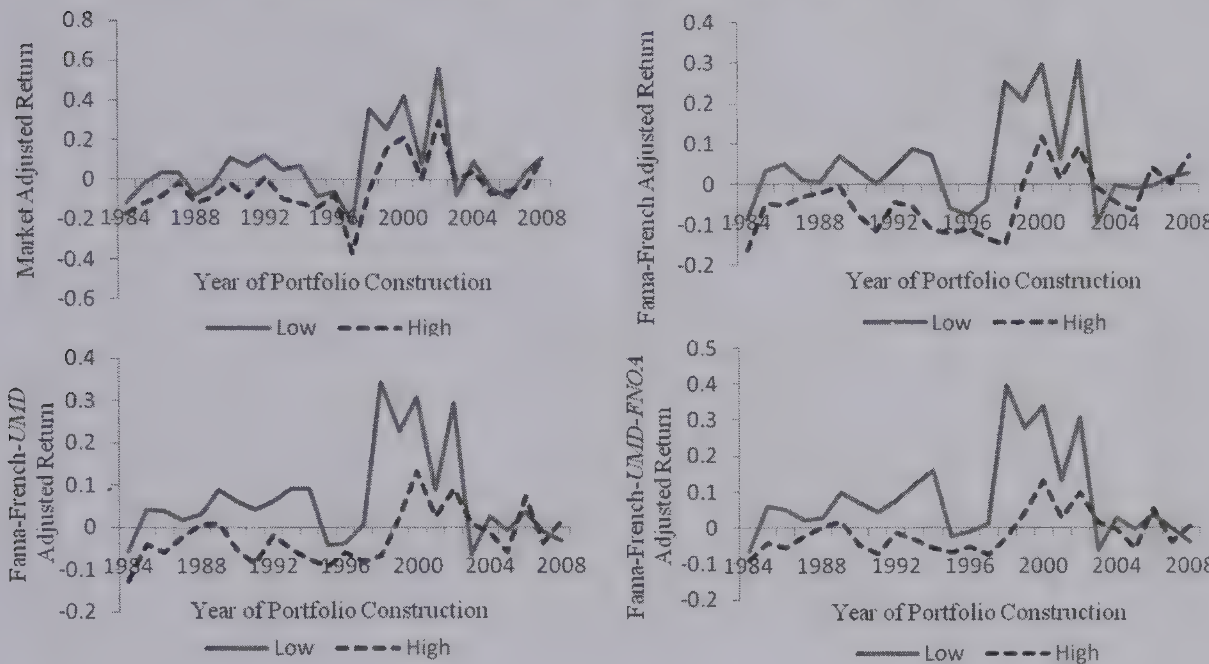
(continued on next page)

TABLE 4 (continued)

Year of Portfolio Construction	Annual Inflation Rate	Market Adjusted Return			Add Factors: Fama-French Adjusted Return			Add *UMD*: Fama-French-*UMD* Adjusted Return			Add *FNOA*: Fama-French-*UMD-FNOA* Adjusted Return		
		Low	High	Hedge	Low	High	Hedge	Low	High	Hedge	Low	High	Hedge
Mean				0.098			0.089			0.088			0.097
# Low > High				23.00			22.00			22.00			22.00
t-statistic				5.05			4.50			4.50			4.65
p-value				<0.001			<0.001			<0.001			<0.001

The table reports mean firm-level returns, calculated over the year subsequent to inflation-based portfolio formation. Specifically, observations with the lowest (highest) *IGL* are sorted annually into portfolio one (ten), where *IGL* is inflation-adjusted earnings minus nominal earnings (scaled), and it captures unrecognized inflation gains and losses. Then, for each portfolio-year, mean abnormal returns are accumulated over the subsequent year beginning three months after the fiscal year-end. Market Adjusted Return is the annually compounded raw return of the firm, $R_{i,t}$, minus the annually compounded value-weighted return on all NYSE, AMEX, and NASDAQ stocks in CRSP. Fama-French (Fama-French-*UMD*, Fama-French-*UMD-FNOA*) Adjusted Returns are $R_{i,t}$ minus the product of firm-specific betas and the respective Fama-French (Fama-French-*UMD*, Fama-French-*UMD-FNOA*) factors, where firm-specific betas are obtained by regressing the firm's time-series monthly excess return on the Fama-French (Fama-French-*UMD*, Fama-French-*UMD-FNOA*) factors. Fama-French factors are *MKTRF*, *SMB*, and *HML*. *UMD* is the Carhart (1997) momentum factor. Raw stock returns are from the CRSP Monthly Stock File, adjusted for delisting returns. The risk-free rate and the monthly factors are from the Fama-French Portfolios and Factors dataset available through WRDS. *NOA* is net operating assets deflated by lagged total assets, constructed as described in Hirshleifer et al. (2004). *FNOA* is a *NOA*-based factor, formed as described in Table 3. The annual inflation rate is the annualized monthly consumer price index obtained from the Global Insight dataset's Basic Economics Monthly Series (PRNEW). p-values are obtained from two-tailed t-tests. The sample includes all NYSE, AMEX, and NASDAQ U.S. stocks that are in the intersection of CRSP and Compustat, that have data available to calculate inflation gains and losses, and with fiscal year-ends between 1984 and 2008 (n = 25).

FIGURE 2
Abnormal Returns from Using Inflation Information: Firm Level



The figure plots the annual mean firm-level returns reported in Table 4. The returns are accumulated over the year subsequent to portfolio formation, for the high and low decile portfolios formed each year based on *IGL*, such that observations with the lowest (highest) *IGL* are sorted into portfolio one (ten). *IGL* is inflation-adjusted earnings minus nominal earnings (scaled), and it captures unrecognized inflation gains and losses.

guishing the monetary and nonmonetary components can explain the abnormal returns findings. Distinguishing monetary and nonmonetary assets is critical in inflationary accounting because inflation affects the two classes of assets differently.

I conduct two tests to investigate this question. In the first test, I begin by using the algorithm to approximate inflation-adjusted earnings estimated without distinguishing monetary and nonmonetary assets (denoted as *IAEarnings_NoDistinguish*), by multiplying all assets by a constant as if there were no difference between the two classes of assets.[23] I then test for a systematic association between future returns and the difference between *IAEarnings_NoDistinguish* and *IAEarnings*—i.e., the inflation-adjusted earnings correctly estimated by distinguishing monetary and nonmonetary assets. Using the relationship between future returns and the expected surprise to infer how investors process information is similar to the approach used in prior studies (e.g., Bernard and Thomas 1990; Sloan 1996).

Table 5 reports the results. Panel A provides inflation-adjusted earnings with and without monetary and nonmonetary assets distinguished across *IGL* portfolios. It shows that not distinguishing monetary and nonmonetary assets leads to overestimation of inflation gains and losses (in

[23] Because of the clean surplus relation, this is equivalent to multiplying the income of all firms by the same constant, which is a function of the annual inflation rate, without taking into account differences in firms' asset composition.

TABLE 5

Future Returns and Investor Processing of Inflation-Adjusted Earnings

	Panel A			Panel B			Panel C			
	Estimate of Current-Period Inflation-Adjusted Earnings, $IAEarnings_t$			Subsequent Abnormal Returns (from Table 3)			Regress α_p on *DIFF*			
IGL Portfolio	*IAEarnings*	*IAEarnings_NoDistinguish*	*DIFF* = *IAEarnings* − *IAEarnings_NoDistinguish*	α_p	t-stat	p-value	Slope	t-stat	p-value	R^2
							0.1507	4.74	0.001	0.738
1	−0.1705	−0.1935	0.0231[a]	0.0049	4.37	<0.001				
2	−0.0205	−0.0249	0.0044	0.0046	4.78	<0.001				
3	0.0094	0.0077	0.0017	0.0042	4.46	<0.001				
4	0.0255	0.0303	−0.0049	0.0041	4.68	<0.001				
5	0.0310	0.0360	−0.0050	0.0030	1.68	0.093				
6	0.0351	0.0415	−0.0065	0.0017	1.62	0.106				
7	0.0385	0.0434	−0.0049	0.0001	0.11	0.913				
8	0.0484	0.0566	−0.0083	−0.0004	−0.38	0.703				
9	0.0744	0.0899	−0.0155	−0.0013	−1.26	0.207				
10	0.1876	0.2308	−0.0432[b]	−0.0036	−3.10	0.002				

[a] Positive news at $t+1$ when higher-than-expected inflation gains turn into cash flows.

[b] Negative news at $t+1$ when lower-than-expected inflation gains turn into cash flows.

The table reports results from processing inflation information and its link to future returns.

Panel A reports current-period inflation-adjusted earnings, scaled by lagged market value of equity. *IAEarnings* refers to inflation-adjusted earnings derived in a manner that distinguishes monetary and nonmonetary assets; this is the *IAEarnings* measure used throughout the study and obtained using the algorithm. *IAEarnings_NoDistinguish* refers to inflation-adjusted earnings obtained using the algorithm without distinguishing monetary and nonmonetary assets.

Panel B reports the Fama-French Alphas, α_p, from Table 3, which are $t+1$ abnormal returns obtained from portfolio regressions constructed on current-period inflation information, i.e., based on $IAEarnings_t$.

Panel C reports results from estimating a regression of α_p on *DIFF* and an intercept (that is omitted), where *DIFF* = *IAEarnings* − *IAEarnings_NoDistinguish*. *IGL* is inflation-adjusted earnings minus nominal earnings (scaled), and it captures unrecognized inflation gains and losses.

p-values are obtained from two-tailed t-tests. The sample includes all NYSE, AMEX, and NASDAQ U.S. stocks that are in the intersection of CRSP and Compustat, that have data available to calculate inflation gains and losses, and with fiscal year-ends between 1984 and 2008 (n = 25).

absolute value), such that *IAEarnings* are overestimated (underestimated) in the low (high) *IGL* portfolio, which results in an increase in the difference between *IAEarnings* and *IAEarnings_NoDistinguish* across *IGL* portfolios. Panel B provides the future abnormal returns (α_p) reported in Table 3 across *IGL* portfolios. Strikingly, the patterns in *IAEarnings* − *IAEarnings_NoDistinguish* (hereafter, *IAEarnings* − *IAEarnings_NoDistinguish* is referred to as *DIFF*) and α_p across the portfolios reveals that the two patterns are closely related, with the highest (lowest) future abnormal return arising when *IAEarnings* is at its highest (lowest) compared to *IAEarnings_NoDistinguish*. Specifically, not distinguishing monetary and nonmonetary assets leads to the greatest underestimation of inflation-adjusted earnings relative to their correctly estimated value (*DIFF* = 0.0231) in the lowest *IGL* portfolio, which is consistent with the return findings in that the lowest *IGL* portfolio has the highest and most significantly positive future abnormal return ($\alpha_p = 0.0049$; $p < 0.001$). Thus, the results provide consistent evidence that the highest future abnormal return is obtained when *IGL* is most underestimated—i.e., when the future realization of inflation gains in cash flows is not fully expected, reflecting a future positive surprise when inflation gains turn into higher-than-expected cash flows. Investigating the highest *IGL* portfolio reveals similar consistency between inflation processing and future abnormal returns, with the lowest differential (*DIFF* = −0.0432) associated with the lowest and most significantly negative future abnormal return ($\alpha_p = -0.0036$; $p = 0.002$). This consistency between the patterns in Panels A and B is also reflected in Panel C, which provides results from regressing α_p on *DIFF* and an intercept that is omitted from the table for brevity. The results show a significantly positive coefficient on *DIFF* (coefficient = 0.1507; p = 0.001).

In sum, the findings from this first test reveal that the error from not fully incorporating the effects of inflation is associated with the documented pattern in future abnormal returns. This test therefore provides evidence that the negative association documented between inflation gains and future abnormal returns is predictable and consistent with an inflation adjustment argument.

In the second test, I use a direct rational expectations framework similar to that employed in Mishkin (1983), Ball and Bartov (1996), and Sloan (1996). Specifically, to examine how inflation information is reflected in stock prices, I estimate the following nonlinear weighted least-squares system of equations:

$$\begin{aligned} CFO_{t+1} &= \delta_0 + \delta_1 \cdot IGL_t + \delta_2 \cdot IGL\_NoDistinguish_t + \delta_3 \cdot CFO_t + \delta_4 \cdot ACCRUALS_t + \zeta_{t+1} Return_{t+1} \\ &= \psi(CFO_{t+1} - \delta_0^* - \delta_1^* \cdot IGL_t - \delta_2^* \cdot IGL\_NoDistinguish_t - \delta_3^* \cdot CFO_t - \delta_4^* \cdot ACCRUALS_t) \\ &\quad + \mu_{t+1}, \end{aligned} \tag{4}$$

where the accounting variables are as defined above, and $Return_{t+1}$ is the adjusted return accumulated over year t+1 and beginning three months after the fiscal year-end of year t. To control for heteroscedasticity, the first equation of the above system is scaled by the ratio of the residual variances of the two equations, where each equation's residual variance is obtained from an estimation of the given equation alone. Following my findings that *IGL*—constructed by distinguishing monetary and nonmonetary assets—predicts future *CFO*, I predict δ_1 (δ_2) to be significantly positive (insignificant). If stock prices correctly reflect the implications of inflation gains for future cash flows, then I expect δ_1^* (δ_2^*) to be significantly positive (insignificant), with the difference between the respective values of these two coefficients not significant. A significant difference between the coefficients on the inflation variables from the two equations would be evidence of investors not fully incorporating the implications of inflation information.

Table 6 reports the results. The table reveals that, as expected, δ_1 is significantly positive ($\delta_1 = 0.0815$; p = 0.010), δ_2 is insignificant ($\delta_2 = -0.0242$; p = 0.289), and the difference between δ_1 and δ_2 is significant (p = 0.044 for the test: $\delta_1 = \delta_2$), consistent with the ability of inflation gains and losses to predict future cash flows. However, the table also reveals that δ_1^* is insignifi-

TABLE 6

Stock Price Reaction to Information in Current Inflation Gains and Losses about Future Cash Flows

$$CFO_{t+1} = \delta_0 + \delta_1 \cdot IGL_t + \delta_2 \cdot IGL\_NoDistinguish_t + \delta_3 \cdot CFO_t + \delta_4 \cdot ACCRUALS_t + \zeta_{t+1}$$

$$Return_{t+1} = \psi(CFO_{t+1} - \delta_0^* - \delta_1^* \cdot IGL_t - \delta_2^* \cdot IGL\_NoDistinguish_t - \delta_3^* \cdot CFO_t - \delta_4^* \cdot ACCRUALS_t) + \mu_{t+1}$$

	Estimate	Asymptotic Standard Error	χ^2-statistic	Asymptotic Pr > χ^2
δ_0	0.0587	0.0009	66.00	<0.001
δ_1	0.0815	0.0315	2.58	0.010
δ_2	−0.0242	0.0228	−1.06	0.289
δ_3	0.6912	0.0057	120.71	<0.001
δ_4	0.0825	0.0056	14.76	<0.001
Ψ	0.3814	0.0130	29.24	<0.001
δ_0^*	−0.2514	0.0131	−19.21	<0.001
δ_1^*	−0.4178	0.2723	−1.53	0.125
δ_2^*	1.6302	0.2044	7.98	<0.001
δ_3^*	0.7215	0.0493	14.62	<0.001
δ_4^*	0.6675	0.0521	12.81	<0.001
Tests				
$\delta_1 = \delta_2$			4.05	0.044
$\delta_1^* = \delta_2^*$			20.01	<0.001
$\delta_1 = \delta_1^*$			10.27	0.001
$\delta_2 = \delta_2^*$			64.71	<0.001

The table reports results from estimating a nonlinear weighted least squares system of equations to examine how inflation information is reflected in stock prices. CFO_{t+1} is cash flows from operating activities for year t+1, extracted directly from the statement of cash flows or derived from changes in balance sheet accounts as explained in Table 1. $ACCRUALS_t$ is total operating accruals, calculated as $ACCRUALS_t = NominalEarnings_t - CFO_t$, where *NominalEarnings* is Income before Extraordinary Items as reported in the financial statements (Compustat: IB). *IGL* captures unrecognized inflation gains and losses and is calculated as $IGL = IAEarnings - NominalEarnings$, where *IAEarnings* is nominal earnings restated to an inflation-adjusted basis using the algorithm in a manner that distinguishes monetary and nonmonetary assets; this is the inflation gains and losses measure used throughout the study. *IGL_NoDistinguish* approximates unrecognized inflation gains and losses without distinguishing monetary and nonmonetary assets and is calculated as $IGL\_NoDistinguish = IAEarnings\_NoDistinguish - NominalEarnings$, where *IAEarnings_NoDistinguish* refers to inflation-adjusted earnings obtained using the algorithm without distinguishing monetary and nonmonetary assets. All accounting variables are scaled. $Return_{t+1}$ is adjusted stock return for year t+1, accumulated over the year subsequent to year t and beginning three months after the fiscal year-end of year t. Raw stock returns are from the CRSP Monthly Stock File, adjusted for delisting returns. To control for heteroscedasticity, the first equation of the system is scaled by a scaling factor computed as the ratio of the residual variances of the two equations, where each equation's residual variance is obtained from estimating the equation alone. p-values are obtained from asymptotic Chi-squared tests. The sample includes all NYSE, AMEX, and NASDAQ U.S. stocks that are in the intersection of CRSP and Compustat, that have data available to calculate inflation gains and losses, and with fiscal year-ends between 1984 and 2008 (n = 25).

cant ($\delta_1^* = -0.4178$; $p = 0.125$), δ_2^* is significantly positive ($\delta_2^* = 1.6302$; $p < 0.001$), the difference between δ_1^* and δ_2^* is significant ($p < 0.001$ for the test: $\delta_1^* = \delta_2^*$), and the differences in the coefficients on the inflation variables across the equations are significant ($p = 0.001$ for the test: $\delta_1 = \delta_1^*$; $p < 0.001$ for the test: $\delta_2 = \delta_2^*$). Taken together, the findings from the second test are consistent with investors not fully incorporating the implications of *IGL* for future *CFO*, and with investors fixating on inflation gains and losses without distinguishing monetary and nonmonetary assets.

In sum, the two tests above use two different but complementary approaches to investigate how investors process inflation information in the cross-section of firms. The findings from these tests provide consistent evidence that the negative return pattern documented across portfolios ofinflation gains is consistent with an inflation adjustment argument, whereby investors do not distinguish monetary and nonmonetary assets.[24]

VII. ROBUSTNESS TESTS

Are the Abnormal Returns Attributable to Risk?

The return analyses provide evidence of mispricing. However, although the abnormal returns can be attributable to mispricing from inflation-adjusted information being costly to obtain and process, they may also be attributable to an omitted inflation-based risk factor. To verify the source of the abnormal returns, I first compare risk characteristics across *IGL* portfolios, and find that there is no pattern in risk characteristics across the portfolios.[25] I also conduct a two-step Fama-MacBeth test, which examines whether a risk factor is priced. This test uses an inflation-based factor and analyzes whether there is a positive risk premium to this factor in a cross-sectional regression analysis (Fama and MacBeth 1973; Fama and French 1992).

In the first step, I estimate time-series regressions at the portfolio level, based on the 25 portfolios constructed on *MVE* and *BTM* (Fama and French 1993):

$$R_{p,m} - R_{f,m} = \lambda_p + \beta_{p,MKTRF,m} \cdot MKTRF_m + \beta_{p,SMB,m} \cdot SMB_m + \beta_{p,HML,m} \cdot HML_m \\ + \beta_{p,UMD,m} \cdot UMD_m + \beta_{p,FIGL,m} \cdot FIGL_m + \kappa_m, \tag{5}$$

where *FIGL* is a monthly factor-mimicking inflation information portfolio that I form based on *IGL* information in a similar way to how *FNOA* is constructed in Section V, following the procedure Fama and French (1993) use in forming *HML* and *SMB*. I focus on the 25 Fama-French portfolios because explaining their cross-sectional pattern in returns has attracted increasing interest in the literature (e.g., Lettau and Ludvigson 2001).[26] I use the time-series portfolio regressions to obtain the predicted factor loadings (betas) for each of the 25 portfolios, estimated using

[24] Note that there may be other channels that contribute to the negative association between *IGL* and future returns. For instance, investors may apply different rules to relate *IGL* to future cash flows and, thus, although future cash flows are higher when inflation gains are higher (Table 2), growth in realized cash flows could be lower than expected. In this case there would be a positive association between *IGL* and future cash flows but a negative relation between *IGL* and abnormal returns. Alternatively, investors may naively suppose that *IGL* turns into future cash flows in a one-to-one manner over some finite horizon, ignoring potentially important differences in the investment process associated with inflation gains. In this case inflation gains may generate future cash flows that are lower than what investors expect, which would lead to negative future returns. In this paper it is difficult to empirically investigate such channels without either a more structural model of the firms' investment opportunities or more finely disaggregated cash flow data that provide details about the source of each firm's cash flows from operating and investment activities; making several additional assumptions would make my inflation measure noisy at best. Importantly, the two tests I use to infer how investors process inflation information are similar to those used in several prior studies.

[25] For example, the median $\beta_{i,MKTRF}$ ($\beta_{i,SMB}$; $\beta_{i,HML}$; $\beta_{i,UMD}$) varies between 1.00 (0.59; 0.11; −0.096) and 1.05 (0.68; 0.46; −0.13), respectively, with no trend across *IGL* portfolios.

[26] More information about the 25 Fama-French portfolios (constructed on five *MVE* and five *BTM* portfolios) is at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

five-year rolling windows that end at month m, with the requirement of at least ten portfolio-month observations in each window.

In the second step, I estimate cross-portfolio monthly regressions of portfolio excess returns for month $m+1$ on the predicted rolling betas, such that each regression pools predicted betas for the 25 portfolios, as follows:

$$R_{p,m+1} - R_{f,m+1} = \phi_{m+1} + \gamma_{MKTRF,m+1} \cdot \hat{\beta}_{p,MKTRF,m} + \gamma_{SMB,m+1} \cdot \hat{\beta}_{p,SMB,m} + \gamma_{HML,m+1} \cdot \hat{\beta}_{p,HML,m} + \gamma_{UMD,m+1} \cdot \hat{\beta}_{p,UMD,m} + \gamma_{FIGL,m+1} \cdot \hat{\beta}_{p,FIGL,m} + \sigma_{m+1}, \quad (6)$$

where $\hat{\beta}_{p,MKTRF,m}$, $\hat{\beta}_{p,SMB,m}$, $\hat{\beta}_{p,HML,m}$, $\hat{\beta}_{p,UMD,m}$ and $\hat{\beta}_{p,FIGL,m}$ are the predicted portfolio betas estimated in the first step using data conditioned on month m. I aggregate and conduct tests on the estimates following the Fama and MacBeth (1973) procedure.

Table 7 reports the results. With respect to the Fama-French factors, the results are consistent with prior research, such as Petkova (2006) and Core et al. (2008). The mean estimated coefficients (p-values associated with the Fama-MacBeth t-statistics) on the market, size, book-to-market, and momentum factors equal −0.0058, 0.0004, 0.0038, and 0.0001 (0.122, 0.857, and 0.033), respectively. Turning to the coefficient of interest, the mean estimated coefficient on $\hat{\beta}_{p,FIGL,m}$ across the monthly cross-sectional regressions, i.e., $\gamma_{FIGL,m+1}$, is −0.0015 with a p-value of 0.318. The insignificance of the coefficient associated with the inflation gains factor, $\gamma_{FIGL,m+1}$, is inconsistent with an omitted risk factor associated with inflation. Rather, this result is consistent with abnormal returns being attributable to mispricing. Such mispricing does not necessarily imply market inefficiency, however, because inflation-adjusted information is potentially costly to obtain and process.

Possible Confounding Factors

In additional robustness tests, I check whether the abnormal return findings can be explained by factors that are known to explain returns in the cross-section. First, nominal leverage and nominal earnings-to-price ratio (E/P) are known to explain average returns. Although all of the return models in this study include as a control the size and book-to-market factors that are shown to absorb the roles of nominal leverage and E/P in explaining average returns (Fama and French 1992), I analyze the pattern in leverage and E/P across the *IGL* portfolios. I find that the respective median leverage and E/P are 0.6 and 0.04 for portfolio one and 0.61 and 0.05 for portfolio ten, with small variation and no systematic pattern in these ratios across portfolios. I also find small variation and no systematic pattern in past returns across the *IGL* portfolios, with a median past 12-month return of 0.10 and 0.12 for portfolios one and ten, respectively. Second, to examine whether industry concentration helps explain the results from the return tests, I examine the percentage of observations from each of the 15 industries in each of the ten inflation gains portfolios. The results suggest that no specific industry dominates, with industry concentration varying from 0.7 percent to 27.3 percent in the most extreme cases across the portfolios. Third, I examine the possible effect of industry-related omitted variables on the positive association between inflation gains and future cash flows by allowing not only the regression constant in Equation (1) to vary by industry, but also the other coefficients. To do so, I estimate Equation (1) by industry (including year fixed effects), and summarize the industry cross-sectional estimates using Fama and MacBeth (1973) t-statistics. The results reveal no change in inferences regarding the significantly positive association between inflation gains and future *CFO*, with cross-industry mean coefficients equal 0.05579, 0.04341, 0.0348, and 0.07375 for the one- through four-year-ahead horizon, respectively.

Fourth, I further examine the robustness of the negative pattern in the return results in Tables

TABLE 7

Two-Step Fama-MacBeth Procedure: Results from Second Step Monthly Cross-Portfolio Regressions

$$R_{p,m+1} - R_{f,m+1} = \phi_{m+1} + \gamma_{MKTRF,M+1} \cdot \hat{\beta}_{p,MKTRF,m} + \gamma_{SMB,m+1} \cdot \hat{\beta}_{p,SMB,m} + \gamma_{HML,m+1} \cdot \hat{\beta}_{p,HML,m} + \gamma_{UMD,m+1} \cdot \hat{\beta}_{p,UMD,m} + \gamma_{FIGL,m+1} \cdot \hat{\beta}_{p,FIGL,m} + \sigma_{m+1}$$

	Prediction	Mean	Std. Dev.	Standard Error	Fama-MacBeth t-statistic	p-value
Φ_{m+1}		0.0125	0.0561	0.0036	3.51	<0.001
$\gamma_{MKTRF,m+1}$		−0.0058	0.0592	0.0038	−1.55	0.122
$\gamma_{SMB,m+1}$		0.0004	0.0345	0.0022	0.18	0.857
$\gamma_{HML,m+1}$	+	0.0038	0.0321	0.0020	1.85	0.033
$\gamma_{UMD,m+1}$		0.0001	0.0611	0.0039	0.03	0.976
$\gamma_{FIGL,m+1}$	?	−0.0015	0.0243	0.0015	−1.00	0.318
Adj. R^2		0.510				

The table presents summary results from monthly cross-portfolio regressions in the second step of a two-step Fama-MacBeth (1973) procedure. In the first stage, portfolio time-series regressions are estimated using the 25 Fama-French portfolios (Fama and French 1993). Value-weighted portfolio excess return for month *m* for each portfolio *p*, $R_{p,m} - R_{f,m}$, is regressed on the contemporaneous monthly returns for the Fama-French (*MKTRF*, *SMB*, *HML*), Carhart (1997) momentum (*UMD*), and inflation-based (*FIGL*) factors, using five-year rolling windows that end at month *m* (requiring at least ten portfolio-month observations in each window), as follows:

$$R_{p,m} - R_{f,m} = \lambda_p + \beta_{p,MKTRF,m} \cdot MKTRF_m + \beta_{p,SMB,m} \cdot SMB_m + \beta_{p,HML,m} \cdot HML_m + \beta_{p,UMD,m} \cdot UMD_m + \beta_{p,FIGL,m} \cdot FIGL_m + \kappa_m.$$

This results in five predicted rolling betas as of month *m*. In the second step, monthly cross-portfolio regressions of portfolio excess returns for month $m+1$, $R_{p,m+1} - R_{f,m+1}$, are regressed on the predicted rolling betas from the first step, such that each regression pools predicted betas for the 25 portfolios estimated in the first step using data conditioned on month *m*. The estimates are aggregated and tested following the Fama and MacBeth (1973) procedure. Fama-French and momentum factors are from the Fama-French Portfolios and Factors dataset available through WRDS. Fama-French 25 portfolio excess returns are from Ken French's website. $R_{f,m}$ is the one-month Treasury bill rate. The inflation-based factor, *FIGL*, is formed following Fama and French (1993), using the same procedure I use to form *FNOA*, as described in Table 3. p-values are obtained from one-tailed (two-tailed) t-tests when there is (no) prediction. The sample includes all NYSE, AMEX, and NASDAQ U.S. stocks that are in the intersection of CRSP and Compustat, that have data available to calculate inflation gains and losses, and with fiscal year-ends between 1984 and 2008 (n = 25).

3–5 by testing whether it is *IGL* and not the balance sheet bloat that causes the return distribution I find across *IGL* portfolios. Specifically, I sort observations into five *NOA* and five *IGL* portfolios (independent sorts) and obtain abnormal returns for the five-by-five, double-sorted portfolios. To assess the pattern across the portfolios, I also estimate a predicted trend regression model by regressing abnormal return on the *IGL* portfolio number (and an intercept). Untabulated results reveal that the negative return pattern across *IGL* portfolios persists in each of the *NOA* portfolios. The trend, which is the slope coefficient from the trend regression, is negative in all *NOA* portfolios (with a trend varies between −0.0056 and −0.0372 across the four abnormal return metrics I use), and significantly so in all cases except three, where the trend is marginally significant (with a p-value varies between 0.001 and 0.104 across the related four abnormal return metrics). These results provide additional evidence that the return distribution in Tables 3–5 is not caused by the balance sheet bloat. I also conduct an analysis wherein I repeat the return tests in Tables 3 and 4

after constructing *FNOA* as explained in Section III, but using ten *NOA* portfolios (instead of two portfolios as in Section III). Thus, *FNOA* is constructed based on the intersection of three *BTM* portfolios and the top and bottom deciles of *NOA*. This is because most of the negative return documented in the balance sheet bloat is concentrated in the top one to two deciles of *NOA*. Untabulated inferences from this analysis are unchanged relative to the findings in Tables 3 and 4. For example, the hedge return in Table 4 is positive in 21 of the 25 sample years, with a mean of 0.097 and $p < 0.001$.[27]

VIII. CONCLUDING REMARKS

I hypothesize and find that inflation information in the current period affects future cash flows. Specifically, I find that unrecognized inflation gains turn into future cash flows from operations over the subsequent four years. I also find that investors do not fully incorporate such information into their investment decisions, in that I find significant abnormal returns for inflation-based trading strategies. I further find that the direction of the abnormal returns is consistent with investors not fully distinguishing monetary and nonmonetary assets. Robustness tests show that the documented abnormal returns are not attributable to differences in risk characteristics, and that an inflation-based factor is not a priced risk factor. These results are consistent with future abnormal returns being attributable to mispricing from costly information rather than to an omitted risk factor. Overall, the findings suggest that inflation has significant implications for performance and stock prices, even when inflation is relatively low.

My findings complement prior studies that primarily investigate stock pricing effects over a short event window or the contemporaneous year. A common theme from these studies is that inflation-adjusted data are inconsequential for making financial decisions. Watts and Zimmerman (1980) and Beaver et al. (1983) suggest that a potential interpretation of this conclusion is that stock prices are not affected by inflation-adjusted data because such information is new, and market participants have not yet learned how to analyze and process it (the "learning effect"). Also, Beaver and Landsman (1983) reason that the inflation-adjusted data are potentially too complex and unfamiliar to use. My study contributes to this literature by indicating that (1) investors do not appear to ignore inflation-adjusted data, which suggests that the learning effect has been partially realized over the past decades, and (2) investors do not fully incorporate inflation information, consistent with such information being complex and unfamiliar. Further, this study is the first to identify the mechanism that leads to the mispricing by linking stock returns to the underlying valuation fundamentals and revealing how inflation-adjusted information is processed by the stock market.

---

[27] I conduct two additional sensitivity tests. First, I estimate a model that pools the ten *IGL* portfolios and allows the intercepts for the extreme portfolios to vary using indicator variables. Second, I examine whether the firm-level return tests hold when using median abnormal returns. Untabulated results reveal unchanged inferences. Also, this study's return findings are distinct from, and cannot be explained by, inflation illusion. The inflation illusion hypothesis (Modigliani and Cohn 1979) posits that highly levered firms are more undervalued because of investors' failure to incorporate the gain accruing from purchasing power depreciation of nominal liabilities. Because the erosion of nominal liabilities leads to higher inflation gains, the direct effect from the inflation illusion hypothesis is higher (lower) future abnormal returns when inflation gains are high (low) as investors who suffer from inflation illusion are positively (negatively) surprised over future periods. However, despite this offsetting effect of the inflation illusion on the findings from the return analyses, my results are incremental to the inflation illusion effect in that I find that future abnormal returns are negatively related to inflation gains (see Section V). The inflation illusion hypothesis also posits that investors irrationally discount inflation-adjusted cash flows using nominal interest rates. Here, in contrast, I investigate how inflation directly affects cash flows, rather than how the cash flows are discounted.

APPENDIX

The appendix develops and validates an algorithm for incorporating inflationary effects into accounting amounts, using only publicly available information, by adjusting nominal to inflation-adjusted amounts on a firm-by-firm basis for a broad sample of firms.[28]

Inflation Adjustment Algorithm

Financial statements can be restated using the balance sheet or the income statement.[29] I rely on the balance sheet to adjust the nominal financial statements.[30]

A. Step 1: Adjustment of Nonmonetary Items

Nonmonetary items are linked to the dollar as of the year-end, but represent either a historical cost or a right (obligation) to receive (deliver) services for which purchasing power is not constant. I adjust these items as follows:

A.1. PPE: I use the PPE life cycle to adjust PPE. An asset's useful life is the period over which the entity expects to consume economic benefits from the asset. Assuming that accounting depreciation, on average, reflects an asset's useful life, the PPE life cycle is the average number of years from the asset's purchase until it is fully depreciated. I thus calculate the PPE life cycle as: $PPELifeCycle_t = (1/n) \cdot \Sigma_{i=t-n+1}[GrossPPE / PPE\ Depreciation]_i$, averaged over the four years prior to year-end t ($n = 4$). Next, I adjust Net *PPE* as follows: $adjNetPPE_t = NetPPE_t \cdot CPI_t / CPI_{t-\tau(t)}$, where adj refers to "adjusted;" t refers to the year t fiscal year-end; $\tau(t)$ is the period prior to fiscal year-end t, stated in annual terms and calculated as $\tau(t) = 0.5 \cdot PPELifeCycle_t$; and *CPI* denotes the Consumer Price Index.[31] If *PPELifeCycle* is negative, missing, or greater than the Compustat median limit of weighted expected useful life among different asset classes, which is calculated based on the expected maximum useful life of different PPE classes (e.g., U.S. Regulation 2003) varying between 20 years (e.g., Machinery and Equipment) and 50 years (e.g., Other Structures and Facilities), then I set it to the median life cycle calculated using the Compustat population over the sample period.[32]

---

[28] An extended Appendix that includes further information, rationale, and illustrative details regarding the algorithm is available from the author upon request.

[29] The clean surplus relation makes the two approaches equivalent. This is because the income statement approach derives inflation-adjusted income before financing expenses by adjusting income statement amounts, whereas the balance sheet approach first calculates inflation-adjusted earnings using two successive balance sheets and then calculates inflation-adjusted financing expenses as the difference between net earnings and income before financing expenses. Inflation-adjusted financing expenses are the same if derived using the balance sheet or the income statement, resulting in same inflation-adjusted earnings under the two approaches.

[30] This is because (1) it avoids mistakes inherent in deriving *IAEarnings* directly from the income statement, (2) it is more accurate because all transaction dates and income statement amounts are not necessary, and (3) because I focus on inflation-adjusted earnings, rather than inflation-adjusted revenues or gross profit, I can bypass reliance on further assumptions necessary to adjust the income statements (e.g., the timing of revenues over the year).

[31] I multiply *PPELifeCycle* by one-half because the life cycle is derived from gross, rather than net, PPE so the expected remaining useful life is one-half the gross PPE life cycle. Information about the exact transaction dates and amounts over the life of the firm is unavailable. Such information could help in estimating the exact purchasing date of each component of PPE and adjust it based on the associated vintage's purchasing power. Instead, I make a simplifying assumption that the PPE in place is acquired evenly over its life with the firm. That is, I adjust PPE using one-half of the Gross PPE life cycle such that the expected value of the remaining useful life is one-half of the life cycle obtained from Gross PPE. Also, note that because the adjustment is accurate to the monthly level, whereas t refers to annual amounts, τ is often a fraction (e.g., for an estimated purchase date of six months prior to fiscal year-end t, $\tau = 0.5$ and $NetPPE_t$ is adjusted using $CPI_t / CPI_{t-0.5}$).

[32] Note that there can be alternative adjustment procedures depending on the assumptions used and the objectives underlying the adjustment. My objectives are to: (1) ensure consistency with actual inflationary GAAP; (2) obtain a sample of firms for which Compustat does not necessarily have available adjustment parameters (e.g., inventory and deprecation methods); and (3) develop a procedure that can be validated on firms in another country. For example, Davidson et al. (1976) requires data on the depreciation method. Requiring data about the inventory and depreciation methods would reduce my sample considerably, because such U.S. data are unavailable for about 40 percent of the observations.

A.2. Inventory: I use the inventory turnover, *IT*, ratio to adjust inventory, using the ratio *COGS/Inventory*, where *COGS* is the Cost of Goods Sold. Year-end t inventory turnover is calculated as: $IT_t = COGS_t / [(INV_t + INV_{t-1}) / 2]$. If, e.g., $IT_t = 2$, the firm invests in inventory twice a year so the average inventory is six months old. In expectation, year t inventory will have a remaining life of $12 / (2 \cdot IT_t)$ when stated in months, or $\kappa(t) = 1 / (2 \cdot IT_t)$ when stated in years. Thus, I adjust inventory as follows: $adjINV_t = INV_t \cdot CPI_t / CPI_{t-\kappa(t)}$. If *COGS* or *INV* are missing or negative, *IT* is set to the median *IT* of the Compustat population over the sample period.

A.3. Intangibles: I calculate the intangibles' remaining life for time t, denoted as $\omega(t)$, as the ratio of intangibles to the amortization of the related intangibles at time t. I assume that, in expectation, the number of years prior to the transaction generating the intangibles equals the remaining years until the amount of intangibles is fully reserved, and thus I adjust intangibles using the price index as of the expected value of the original transaction date, or $adjIntangibles_t = Intangibles_t \cdot CPI_t / CPI_{t-\omega(t)}$. I set intangibles' remaining life to the median remaining life of intangibles for the Compustat population over the sample period if it is negative, missing, or greater than firms' common weighted useful life of different intangibles classes, which is calculated based on the useful life of different intangibles classes varying between 2 and 40 years (e.g., patents) and between 20 and 40 years (e.g., goodwill). Also, according to SFAS 142 (effective in 2002), goodwill and other intangible assets no longer have a defined life for amortization but instead are tested annually for impairment. Because the algorithm uses amortization based on the pre-SFAS 141/142 period, it uses parameters obtained from the Compustat population to adjust the years that follow. I repeat all analyses without amortizing the years subsequent to 2002, and the inferences are unchanged.

A.4. Common Stock, Preferred Stock, and Capital Surplus: These items, which are included in shareholders' equity and represent purchasing power as of the stock issue dates, consist of two layers: (1) all stock issues from a firm's establishment through $t-1$, and (2) new equity issues occurring in year t (this layer can include several sub-layers, one from every equity issue that occurred over the year). I assume that equity issues are distributed uniformly over the year. To state amounts in constant dollars as of the reporting date, I begin by adjusting the first layer to derive retained earnings for both year $t-1$ and year t. In constant dollars as of t year-end, the adjusted amount of the first layer in $t-1$ is equal to the amount in t for calculating year t adjusted earnings. Using this two-layer process allows one to adjust earnings without having information about all the preferred and common stock issue dates and amounts from firms' incorporation dates until $t-1$. Thus, the following amount, which corresponds to the first layer and provides $t-1$ equity, appears in any two consecutive retained earnings and is used to extract inflation-adjusted earnings: $adjE_{t-1} = [CommonStock + PreferredStock + CapitalSurplus]_{t-1} \cdot CPI_t / CPI_{t-1}$. For the second layer, I obtain adjusted new issues during the year, $AdjNewIssues_t$, by calculating new issues, $NewIssues_t = [CommonStock + PreferredStock + CapitalSurplus]_t - [CommonStock + PreferredStock + CapitalSurplus]_{t-1}$, and adjusting this amount using one-half year's change in *CPI*, under the assumption that new issues occur uniformly throughout the year.

A.5. Other Monetary Items in Stockholders' Equity but not in Retained Earnings (O): Because earnings are obtained from the difference in retained earnings between two successive periods (adjusted for dividends and capital changes), it is necessary to exclude items that violate the clean surplus relation (e.g., Employee Benefit Trust) from inflation-adjusted retained earnings. This component is assumed to be monetary and is calculated as $O_t = TotalAssets_t - TotalLiabilities_t - ReExOCI_t - CommonStock_t - PreferredStock_t - CapitalSurplus_t$, where $ReExOCI_t$ is per A.6 below.

A.6. Retained Earnings Excluding Other Comprehensive Income (*ReExOCI*): It is critical to maintain the clean surplus relation when deriving earnings. Accordingly, I obtain nominal and

inflation-adjusted Retained Earnings Excluding Other Comprehensive Income. The inflation-adjusted amount is required because *IAEarnings* is derived using the two-period difference in inflation-adjusted *ReExOCI*. The nominal amount is used to derive *O* (per A.5.) as follows: *ReExOCI* = Retained Earnings (Compustat: RE) − Accumulated Other Comprehensive Income (Compustat: ACOMINC). The inflation-adjusted *ReExOCI* as of year *t*, $adjReExOCI_t$, is derived by using the relation that total assets equal total liabilities plus shareholders' equity, and by stating all balance sheets amounts in constant dollars, where monetary (nonmonetary) items are not (are) adjusted: $adjReExOCI_t = adjINV_t + adjNetPPE_t + adjIntangibles_t + OA_t - adjE_{t-1} - adjNewIssues_t - O_t - TotalLiabilities_t$ (where, as above, $adjE_{t-1}=[CommonStock+PreferredStock+Capital\ Surplus]_{t-1} \cdot CPI_t/CPI_{t-1}$). Total liabilities are treated as monetary. I treat as monetary other assets (*OA*) that are not directly adjusted, and derive them as a residual value, using the relation that total assets equal total liabilities plus shareholders' equity, as follows: $OA_t = TotalAssets_t - INV_t - NetPPE_t - Intangibles_t$.

A.7. Other Comprehensive Income and Other Items Affecting Retained Earnings without Directly Affecting Net Income (*OtherInReExOCI*): This item is used in the equation that derives *IAEarnings*. Two types of exclusions are subtle, yet necessary for the accounting identities to hold and thus for the accuracy of the algorithm. First, because *IAEarnings* is obtained using the two-period difference in *adjReExOCI*, dividends must be included in the adjustment. Second, all transactions that are neither part of Other Comprehensive Income nor part of Net Income need to be excluded (e.g., Net Issues of Common Stock under Employee Plans; Purchases and Sales of Treasury Stocks under Employee Plans). Because these exclusions are the result of transactions occurring at the year-end, I treat them as monetary. These amounts are calculated as: $OtherInReExOCI_t = ReExOCI_t - ReExOCI_{t-1} - NetIncome_t + CommonDividends_t + PreferredDividends_t$.

A.8. Dividends: Because dividends are usually paid quarterly, the adjusted common and preferred dividends, *adjCommonDividends* and *adjPreferredDividends*, are adjusted assuming these payments are distributed uniformly over the year.

B. Step 2: Treatment of Monetary Items

Monetary assets and liabilities are measured on the basis of a fixed number of dollars required for their settlement. Thus, nominal monetary amounts are already stated in terms of constant purchasing power and, accordingly, I treat monetary items as equal to their recognized nominal amounts. The following are considered monetary: Cash, Short-Term Investments, Total Receivables, Total Liabilities, and assets not directly treated as nonmonetary assets (*OA*). The inclusion of *OA* implicitly treats unconsolidated but wholly-owned subsidiaries as monetary, consistent with Bernard and Hayn (1986).

C. Final Step: Derivation of Inflation-Adjusted Earnings

Inflation-adjusted earnings, *IAEarnings*, are calculated as follows:

$$IAEarnings_t = [adjReExOCI_t - adjReExOCI_{t-1}] + adjCommonDividends_t + adjPreferredDividends_t - OtherInReExOCI_t - adjExtraordinaryItems_t.$$

I obtain $adjReExOCI_{t-1}$ analogously to $adjReExOCI_t$ (see A.6. above), except that in this case (1) I adjust the accounting amounts reported for year $t-1$ to the purchasing power as of t year-end, and (2) I do not subtract $adjNewIssues_{t-1}$ because it is already part of $adjE_{t-1}$ as the new issues

during $t-1$ are part of the $t-1$ equity amount.[33] To reduce measurement error from the adjustment procedure, I delete observations each year in the top and bottom percentiles of the MVE_{t-1} deflated difference between *IAEarnings* and *NominalEarnings*. Because I investigate the behavior of *IAEarnings* versus *NominalEarnings* and because *NominalEarnings* refers to Net Income Excluding Extraordinary Items, I exclude extraordinary items when deriving *IAEarnings* to make the two earnings measures comparable. I assume that extraordinary items, if any occur, are distributed uniformly over the year and thus are adjusted using one-half year's change in the price index; these items are denoted as $adjExtraordinaryItems_t$.[34]

External Validation of the Algorithm

To provide evidence on the external validity of the algorithm, I test the algorithm on a sample of Israeli firms. Until 2003 Israeli firms were required to recognize financial statements in inflation-adjusted terms and disclose in footnotes the same financial statements in nominal terms, and similar to the U.S., the inflation rate in Israel over the past decade was relatively low. In the validation analysis, I examine the extent to which nominal earnings derived by the algorithm, $NominalEarnings^{Model}$, approximates disclosed nominal earnings, $NominalEarnings^{Actual}$, by estimating the equation: $NominalEarnings^{Model} = \alpha + \beta \cdot NominalEarnings^{Actual} + \varepsilon$. If the algorithm does a good job translating earnings from one measurement basis into the other, I predict the intercept to be equal to 0 and the slope to be equal to 1. Thus, I conduct the tests: $H_0^{\alpha}: \alpha = 0$ against $H_1^{\alpha}: \alpha \neq 0$, and $H_0^{\beta}: \beta = 1$ against $H_1^{\beta}: \beta \neq 1$. To do so, I hand-collect data from Israeli firms' annual nominal and inflation-adjusted financial statements over the 1995–2003 period for 81 randomly selected firms listed on the Tel-Aviv 100 index. This index comprises the 100 firms with the highest *MVE* and accounts for more than 80 percent of the total market's capitalization. The 81 firms that I sample account for 86.63 percent of this index's total market capitalization as of December 21, 2005.

After implementing the algorithm and requiring the same restrictions as with the U.S. data, the inflation-adjusted Israeli sample includes 503 firm-year observations. Also, because footnotes are not always attached to the financial statements, causing nominal footnote disclosures to not always be available, I randomly select 50 firms and gather nominal information, when such footnotes are available. Monthly CPI and exchange rate data are obtained from the Israeli Central Bureau of Statistics. The Israeli sample reflects a median firm size of $220 million. The mean and median values of the difference between actual (i.e., reported) inflation-adjusted earnings and

---

[33] Specifically, $adjReExOCI_{t-1} = adjINV_{t-1} + adjNetPPE_{t-1} + adjIntangibles_{t-1} + adjOA_{t-1} - adjE_{t-1} - adjO_{t-1} - adjTotalLiabilities_{t-1}$, where: $adjINV_{t-1} = INV_{t-1} \cdot CPI_t / CPI_{t-1-\kappa(t-1)}$; $adjNetPPE_{t-1} = NetPPE_{t-1} \cdot CPI_t / CPI_{t-1-\tau(t-1)}$; $adjIntangibles_{t-1} = Intangibles_{t-1} \cdot CPI_t / CPI_{t-1-\omega(t-1)}$; $adjOA_{t-1} = OA_{t-1} \cdot CPI_t / CPI_{t-1}$; $adjO_{t-1} = O_{t-1} \cdot CPI_t / CPI_{t-1}$; as above, $adjE_{t-1} = [CommonStock + PreferredStock + CapitalSurplus]_{t-1} \cdot CPI_t / CPI_{t-1}$; $adjTotalLiabilities_{t-1} = TotalLiabilities_{t-1} \cdot CPI_t / CPI_{t-1}$; and $\kappa(t-1)$, $\tau(t-1)$, and $\omega(t-1)$ refer to the period (stated in years) from which the lagged nonmonetary assets *INV*, *NetPPE*, and *Intangibles*, respectively, are adjusted.

[34] With respect to the derivation of *IGL*, there is a normalization based on a reference point underlying the adjustment procedure. Specifically, accounting amounts can be adjusted to be stated based on either constant dollars to maintain transactions in purchasing power, or current dollars to maintain transactions in consumption units. In the cross-section, the variation in *IGL*, rather than its level, is informative for explaining variation across firms, and the two approaches are equivalent when intercepts are added to the tests. I choose to adjust for constant dollars, leading *IGL* to be more frequently negative. Alternatively, *IGL* can be adjusted such that it is more frequently positive but the variation across firms and over time is unchanged. Accordingly, if the prediction model is $CF_{t+1} = a + b \cdot IGL_t + X_t + \eta_{t+1}$, where X is a vector of additional explanatory variables (conditioned on the time t information set), analyses throughout the study pertain to the parameter b, which is invariant to the reference point underlying the measurement system. The intercept, a, varies with the measurement system but is not a parameter of interest in my prediction analyses. Accordingly, the research design throughout my study includes intercepts in all cross-sectional tests and focuses on the coefficient on *IGL*.

nominal earnings, IGL^{Actual}, are −0.02 and −0.01, with a standard deviation of 0.07. This suggests a difference of about 1 to 2 percent of firms' size, with large variation between the two measures.[35]

The results reveal that the null hypotheses of $\alpha = 0$ (p = 0.609) and $\beta = 1$ (p = 0.240) cannot be rejected, with point estimates of $\alpha = 0.01$ and $\beta = 0.8$.[36] Overall, although the adjustment procedure does not use data about the timing and amounts of all of the firms' transactions over the life of the firms until the reporting date (which are needed for complete inflation adjustment), the findings reveal that the algorithm provides a reasonable and unbiased proxy for the effects of inflation.

REFERENCES

Aboody, D., M. E. Barth, and R. Kasznik. 1999. Revaluations of fixed assets and future firm performance: Evidence from the UK. *Journal of Accounting and Economics* 26 (1–3): 149–178.

Ball, R. 1994. The development, accomplishments and limitations of the theory of stock market efficiency. *Managerial Finance* 20 (2–3): 3–48.

———, and E. Bartov. 1996. How naive is the stock market's use of earnings information? *Journal of Accounting and Economics* 21 (3): 319–337.

Bar-Yosef, S., and B. Lev. 1983. Historical cost earnings versus inflation-adjusted earnings in the dividend decision. *Financial Analysts Journal* 39 (2): 41–50.

Barth, M. E. 1994. Fair value accounting: Evidence from investment securities and the market valuation of banks. *The Accounting Review* 69 (1): 1–25.

———, W. H. Beaver, J. R. M. Hand, and W. R. Landsman. 1999. Accruals, cash flows, and equity values. *Review of Accounting Studies* 4 (3–4): 205–229.

———, D. P. Cram, and K. K. Nelson. 2001. Accruals and the prediction of future cash flows. *The Accounting Review* 76 (1): 27–58.

———, Y. Konchitchki, and W. R. Landsman. 2010. Cost of capital and earnings transparency. Working paper, Stanford University.

Basu, S., S. Markov, and L. Shivakumar. 2006. Systematic errors in analysts' forecasts: Evidence from analysts' use of inflation information. Working paper, Emory University and London Business School.

Beaver, W. H. 1979. Accounting for inflation in an efficient market. In *The Impact of Inflation on Accounting: A Global View. The International Journal of Accounting* 21–42. London, U.K.: Elsevier.

———. 1981. Market efficiency. *The Accounting Review* 56 (1): 23–37.

———, A. A. Christie, and P. A. Griffin. 1980. The information content of SEC Accounting Series Release No. 190. *Journal of Accounting and Economics* 2 (2): 127–157.

———, P. A. Griffin, and W. R. Landsman. 1983. How well does replacement cost income explain stock return? *Financial Analysts Journal* 39 (2): 26–30.

---

35 The validation analysis requires that several obstacles be overcome. First, because the requisite Israeli data are not available in organized format, I hand-collect firms' annual data, as described above. Second, because Israeli GAAP requires footnote disclosure of selected nominal data, considerably more data are reported on an inflation-adjusted basis. Thus, I use an inverted algorithm that maps from inflation-adjusted to nominal amounts, and use as input the Israeli inflation-adjusted data. Third, I create a translation dictionary that classifies different accounting terms with the same content under a specific term and matches each Israeli data item to the equivalent Compustat item. This procedure results in Israeli firm-year observations with a format similar to that of U.S. companies in Compustat.

36 I conduct further checks on the algorithm's accuracy. First, I form a statistic based on the mean difference between reported nominal earnings and earnings obtained from the algorithm, denoted as μ_x, and test $H_0: \mu_x = 0$ against $H_1: \mu_x \neq 0$. The results show that the null cannot be rejected (p = 0.744), which suggests the algorithm provides a reasonable estimate of the effects of inflation. Second, the algorithm uses computations that interact accounting items with monthly CPI values. To investigate whether these computations introduce measurement error, I derive *IAEarnings* after injecting a constant zero inflation rate into the system. This check results in *IAEarnings* being equal to *NominalEarnings*, consistent with zero inflation and zero measurement error from CPI computations. Third, I derive *NominalEarnings* using the algorithm and compare it to the Compustat amount. The results show the same earnings amount in all observations except those with missing values because of unavailable data.

——, M. McNichols, and R. Price. 2007. Delisting returns and their effect on accounting-based market anomalies. *Journal of Accounting and Economics* 43 (2–3): 341–368.

——, and W. R. Landsman. 1983. *Incremental Information Content of Statement 33 Disclosures*. FASB Research Report. Norwalk, CT: FASB.

Bernard, V. L. 1986. Unanticipated inflation and the value of the firm. *Journal of Financial Economics* 15 (3): 285–321.

——, and C. Hayn. 1986. Inflation and the distribution of the corporate income tax burden. *National Tax Journal* 39 (2): 171–187.

——, and J. K. Thomas. 1990. Evidence that stock prices do not fully reflect the implications of current earnings for future earnings. *Journal of Accounting and Economics* 13 (4): 305–340.

Campbell, J. Y., and T. Vuolteenaho. 2004. Inflation illusion and stock prices. *The American Economic Review* 94 (2): 19–23.

Carhart, M. M. 1997. On persistence in mutual fund performance. *The Journal of Finance* 52 (1): 57–82.

Chordia, T., and L. Shivakumar. 2005. Inflation-illusion and post-earnings-announcement drift. *Journal of Accounting Research* 43 (4): 521–556.

Core, J. E., W. R. Guay, and R. Verdi. 2008. Is accruals quality a priced risk factor? *Journal of Accounting and Economics* 46 (1): 2–22.

Davidson, S., C. P. Stickney, and R. L. Weil. 1976. *Inflation Accounting: A Guide for the Accountant and the Financial Analyst*. New York, NY: McGraw-Hill.

Davis-Friday, P. Y., and E. A. Gordon. 2005. Relative valuation roles of equity book value, net income, and cash flows during a macroeconomic shock: the case of Mexico and the 1994 currency crisis. *Journal of International Accounting Research* 4 (1): 1–21.

Dechow, P. M., R. G. Sloan, and A. Sweeney. 1995. Detecting earnings management. *The Accounting Review* 70 (2): 193–225.

Easman, W., A. Falkenstein, and R. Weil. 1979. The correlation between sustainable income and stock returns. *Financial Analysts Journal* 35 (5): 44–48.

Fama, E. F., and K. R. French. 1992. The cross-section of expected stock returns. *The Journal of Finance* 47 (2): 427–465.

——, and ——. 1993. Common risk factors in the returns of stocks and bonds. *Journal of Financial Economics* 33 (1): 3–56.

——, and J. D. MacBeth. 1973. Risk, return, and equilibrium: Empirical tests. *The Journal of Political Economy* 81 (3): 607–636.

——, and G. W. Schwert. 1977. Asset returns and inflation. *Journal of Financial Economics* 5 (2): 115–146.

Financial Accounting Standards Board (FASB). 1978. *Objectives of Financial Reporting by Business Enterprises*. Statement of Financial Accounting Concepts No. 1. Norwalk, CT: FASB.

——. 1984. *Recognition and Measurement in Financial Statements of Business Enterprises*. Statement of Financial Accounting Concepts No. 5. Norwalk, CT: FASB.

Gheyara, K., and J. Boatsman. 1980. Market reaction to the 1976 replacement cost disclosures. *Journal of Accounting and Economics* 2 (2): 107–125.

Gordon, E. A. 2001. Accounting for changing prices: The value relevance of historical cost, price level, and replacement cost accounting in Mexico. *Journal of Accounting Research* 39 (1): 177–200.

Gow, I. D., G. Ormazabal, and D. J. Taylor. 2010. Correcting for cross-sectional and time-series dependence in accounting research. *The Accounting Review* 85 (2): 483–512.

Grossman, S. J., and J. E. Stiglitz. 1980. On the impossibility of informationally efficient markets. *The American Economic Review* 70 (3): 393–408.

Hackel, K. S., J. Livnat, and A. Rai. 2000. A free cash flow investment anomaly. *Journal of Accounting, Auditing & Finance* 15 (1): 1–24.

Hirshleifer, D., K. Hou, S. H. Teoh, and Y. Zhang. 2004. Do investors overvalue firms with bloated balance sheets? *Journal of Accounting and Economics* 38: 297–331.

Kothari, S. P., and J. Shanken. 2004. Asset allocation with inflation-protected bonds. *Financial Analysts Journal* 60 (1): 54–70.

Lettau, M., and S. Ludvigson. 2001. Resurrecting the (C)CAPM: A cross-sectional test when risk premia are time-varying. *The Journal of Political Economy* 109 (6): 1238–1287.

Mishkin, F. S. 1983. *A Rational Expectations Approach to Macroeconomics: Testing Policy Ineffectiveness and Efficient-Market Models*. Chicago, IL: University of Chicago Press for the National Bureau of Economic Research.

Modigliani, F., and R. A. Cohn. 1979. Inflation, rational valuation, and the market. *Financial Analysts Journal* 35 (2): 24–44.

Petersen, M. A. 2009. Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies* 22 (1): 435–480.

Petkova, R. 2006. Do the Fama-French factors proxy for innovations in predictive variables? *The Journal of Finance* 61 (2): 581–612.

Phelps, E. S. 1967. Phillips curves, expectations of inflation, and optimal unemployment over time. *Economica* 34 (135): 254–282.

Ritter, J. R., and R. S. Warr. 2002. The decline of inflation and the bull market of 1982–1999. *Journal of Financial and Quantitative Analysis* 37 (1): 29–61.

Ro, B. T. 1980. The adjustment of security returns to the disclosure of replacement cost accounting information. *Journal of Accounting and Economics* 2 (2): 159–189.

Shumway, T., and V. A. Warther. 1999. The delisting bias in CRSP's Nasdaq data and its implications for the size effect. *The Journal of Finance* 54 (6): 2361–2379.

Sloan, R. G. 1996. Do stock prices fully reflect information in accruals and cash flows about future earnings? *The Accounting Review* 71 (3): 289–315.

Summers, L. H. 1981. Inflation and the valuation of corporate equities. Working Paper No. 824. Washington, D.C.: National Bureau of Economics Research.

Unites States Regulations. 2003. *Property, Plant, and Equipment Departmental Regulation*. Washington, D.C.: Office of the Chief Financial Officer.

Watts, R. L., and J. L. Zimmerman. 1980. On the irrelevance of replacement cost disclosures for security prices. *Journal of Accounting and Economics* 2 (2): 95–106.

THE ACCOUNTING REVIEW
Vol. 86, No. 3
2011
pp. 1087–1115

American Accounting Association
DOI: 10.2308/accr.00000043

The Effect of Annual Report Readability on Analyst Following and the Properties of Their Earnings Forecasts

Reuven Lehavy
Feng Li
Kenneth Merkley
University of Michigan

ABSTRACT: This study examines the effect of the readability of firms' written communication on the behavior of sell-side financial analysts. Using a measure of the readability of corporate 10-K filings, we document that analyst following, the amount of effort incurred to generate their reports, and the informativeness of their reports are greater for firms with less readable 10-Ks. Additionally, we find that less readable 10-Ks are associated with greater dispersion, lower accuracy, and greater overall uncertainty in analyst earnings forecasts. Overall, our results are consistent with the prediction of an increasing demand for analyst services for firms with less readable communication and a greater collective effort by analysts for firms with less readable disclosures. Our results contribute to the understanding of the role of analysts as information intermediaries for investors and the effect of the complexity of written financial communication on the usefulness of this information.

Keywords: *financial analysts; readability; Gunning Fog Index; analyst following.*

Data Availability: *The data used in this study are available from the sources indicated in the text.*

I. INTRODUCTION

Over the past two decades, changes in financial and reporting regulations (e.g., changes in segment disclosures, employee stock option reporting, and Sarbanes-Oxley disclosures) have significantly increased the amount of required disclosures by firms to external users. In addition, technological advancement and new developments in financial engineering have made it more challenging for firms to communicate information about the underlying fundamentals of their businesses in a clear and informative manner. The increase in the amount of required dis-

We thank Beverly Walther (editor), Steven Kachelmeier (senior editor), and two anonymous reviewers. We also thank workshop participants from The University of Chicago, University of Michigan, University of Minnesota, Erasmus University, University of Melbourne, Tel Aviv University, the 2010 AAA Annual Meeting, the 2010 EEA Annual Congress, and a joint presentation of HEC-Paris and INSEAD for their helpful comments and suggestions.

Editor's note: Accepted by Beverly Walther.

Submitted: December 2009
Accepted: September 2010
Published Online: April 2011

closure accompanied by the challenging task of communicating increasingly complex business transactions to investors has led to concerns about the effectiveness of management communication and the ability of interested users to make informed decisions based on this information. As a result, both the SEC and the popular press have routinely criticized firms for the complexity of the language in their 10-K filings (Schroeder 2002). These groups have expressed concerns about the ability of financial statement users, especially small investors, to understand the complicated writing found in firms' financial documents (SEC 1998; Schroeder 2002; Cox 2007).

Given the increasing complexity of firm disclosures and the related concerns about their usability, a natural question arises as to the role of financial analysts in intermediating such information to investors. Specifically, do financial analysts use their expertise to examine this complex communication and provide useful information to financial statement users, or do they prefer to focus their efforts on firms with less complex communication? We attempt to answer this question by examining how the behavior of financial analysts relates to the readability of firms' written communication.

Prior literature examines the relation between the properties of firms' financial disclosures and the behavior of security analysts. These properties include the informativeness of disclosures (Lang and Lundholm 1996; Healy et al. 1999), the use of segment disclosures (Botosan and Harris 2000), and the effect of intangible assets (Barth et al. 2001). These studies generally document that firms with better disclosure quality tend to attract greater analyst following. Other studies examine the effect of certain financial reporting items on the properties of analyst earnings forecasts. These studies find that analyst forecasts are influenced by firms' accounting choices (Hopkins et al. 2000; Bradshaw et al. 2008), changes in the tax law (Plumlee 2003), the clarity of the income effects of specific financial items (Hirst and Hopkins 1998; Hirst et al. 2003), and international diversification (Duru and Reeb 2002).

While each of these studies contributes to our understanding of the effect of firms' financial reporting and disclosure choices on analyst behavior, they generally focus on the effect of a single financial statement item (e.g., interest rates, taxes) or disclosure (e.g., segment reporting).[1] This approach, however, does not explicitly incorporate the potential effect of the properties of other financial statement items or disclosures on analyst following or the properties of their earnings forecasts. More important, this approach does not address the potential impact on analyst behavior of the *overall* complexity of firms' financial communication, that is, the cost of processing and interpreting the entirety of firms' disclosures. This potential limitation is particularly important, given the broad nature of corporate disclosures such as 10-K filings that describe a significant number of interrelated financial items.

We attempt to address this issue by examining the relation between a comprehensive measure of the overall readability of corporate 10-K filings and analyst behavior. This measure, known as the Gunning Fog Index (hereafter, Fog Index), incorporates the number of words per sentence and the number of complex words in a document to derive a measure of the readability or syntactic complexity of firms' 10-K filings (Jones and Shoemaker 1994; Courtis 1995; Li 2008).[2] The Fog Index has been widely used in social science research for several decades to examine the relation

[1] One notable exception is Lang and Lundholm (1996) who use AIMR scores as a measure of overall disclosure quality. The potential limitations of this measure, however, are that it is based on a subjective survey of analysts, limited to a subset of large firms, and no longer available after 1995.

[2] Specifically, we focus on the readability of firms' written communication as a measure of the syntactic complexity of such communication. Stimulated by Chomsky's (1965) seminal generative grammar, there has been a great deal of theoretical and empirical work on what constitutes syntactic complexity and how it functions as a determinant of both sentence comprehension (Gibson 1998) and sentence production (Thompson and Faroqi-Shah 2002). According to Stone et al. (2005, 341–343), sentence complexity is affected by the features of open-class words (nouns and verbs) and their relationships and the number and type of syntactic operations.

between the readability of written information and various decisions or outcomes.[3] Further, recent studies use the Fog Index to examine the readability of annual reports in connection with earnings persistence (Li 2008), timely price adjustment (Callen et al. 2009), and investment efficiency (Biddle et al. 2009).

The Fog Index offers several important advantages. First, it allows us to study a large and diverse group of firms. Second, it is an objective measure, not based on analyst surveys or opinion, and can be calculated for any narrative disclosure. Finally, it allows us not only to capture the effect of the readability of a variety of financial items, but, more important, to directly examine the overall syntactic complexity of firms' written public communication, over and above its specific content.[4]

We focus on the readability of the 10-K filing for several reasons. First, 10-K filings are required for all publicly traded companies and are frequently cited as an input in the decisions of investors and financial analysts (Previts et al. 1994; Rogers and Grant 1997). Second, both the SEC and popular press have criticized firms for the complexity of their language in these filings and the effect of this complexity on investors (Schroeder 2002). You and Zhang (2009) provide evidence that suggests that investors underreact to the information provided in 10-K filings and that this effect is stronger for firms with more complex 10-K reports. Finally, the 10-K filing contains a significant amount of written communication or narrative to use in interpreting its readability. These reasons make the 10-K filing an interesting setting to examine the effects of the readability of firm communication on the activities of financial analysts.

We test the relation between the readability of 10-K filings and several measures of analyst behavior, including analyst following, analyst forecast revision response time, the information content of analysts' reports, and the properties of analyst earnings forecasts. We begin our analysis by examining the relation between analyst following and 10-K filing readability. We argue that, on the one hand, lower readability of firm financial disclosures increases the cost of processing the information in these disclosures and therefore will increase the demand for analyst services. If the cost of obtaining analyst reports is less than the information-processing costs of firm disclosures, then analyst following should be greater for firms with less readable 10-K filings. On the other hand, less readable disclosure could also increase the costs of analyst coverage. That is, analysts could bear greater information-processing costs and higher private search costs, which could lead to less accurate forecasts. If these costs are significant, then analyst following should be lower for firms with less readable disclosures.

Consistent with the prediction of an increasing demand for analyst services for firms with less readable 10-K filings, we document a positive and significant association between a firm's 10-K Fog Index and the number of analysts who cover the firm, even after controlling for other factors related to analyst coverage. We interpret this result as being consistent with a greater collective effort by analysts for firms with less readable disclosures.

Next, we examine whether individual analyst effort is associated with the readability of the 10-K filing. We measure individual analyst effort as the length of time required for analysts to issue their first forecast revision following the 10-K filing. If analysts bear costs in following firms

---

[3] The Fog Index has been used in a variety of applications including medical error, consumer drug use, consumer warranties, mutual fund prospectuses, jury instructions, and academic research prestige. For specific examples, see Charrow and Charrow (1979), Armstrong (1980), Shuptrine and Moore (1980), Ott and Hardie (1997), Gazmararian et al. (1999), Koo et al. (2003), and Lee et al. (2006).

[4] In a recent speech, former SEC Chairman Christopher Cox specifically identifies readability measures as a tool to examine communication complexity. He states, "Just as the Black-Scholes model is commonplace when it comes to compliance with the stock option compensation rules, we may soon be looking to the Gunning-Fog and Flesch-Kincaid models to judge the level of compliance with the plain English rules" (Cox 2007). Similarly, Core (2001) proposes the use of computational linguistics methods in accounting to measure corporate disclosure quality.

with less readable disclosures, then it should take them more time, on average, to issue reports following less readable disclosures. Using a variety of empirical approaches, we find that analysts who cover firms with less readable 10-K filings take a longer time to issue their first report following the 10-K filing. This evidence is consistent with the argument that analysts exert more effort to follow firms with less readable disclosures.

We then turn to an examination of the relation between 10-K readability and the information content of analyst reports. Similar to Frankel et al. (2006), we measure information content as the proportion of a firm's stock returns related to analyst forecast revisions to the total firm's stock return during the time period between the 10-K filing and the subsequent fiscal year-end. We predict that, if investors place greater reliance on analyst reports issued for firms with less readable 10-K filings (potentially due to the higher costs associated with processing and interpreting a complex report), then the information content of analyst reports, or the proportion of firm returns associated with them, should be higher for firms with less readable disclosures. Consistent with this prediction, we find evidence that the proportion of firm returns associated with analysts' reports is higher for firms with less readable 10-K reports. This evidence suggests that investors find analysts' reports more informative for firms with less readable corporate disclosures.

Finally, we examine how 10-K filing readability relates to commonly studied properties of analyst earnings forecasts: dispersion, accuracy, and two measures of earnings consensus forecast uncertainty (overall and common uncertainty) based on Barron et al. (1998). Prior literature finds that the 10-K filing represents a major source of information used by analysts (Previts et al. 1994; Rogers and Grant 1997). However, studies also find that the complexity of items in the 10-K filing affects analysts' use of this information. For example, Plumlee (2003) finds that analyst forecasts are less accurate if they are associated with complex changes in the tax law. Bradshaw et al. (2008) find that differences in accounting choice negatively affect forecast accuracy and increase dispersion. Our tests relate to these findings, because we examine the effect of complexity through the ability of analysts to incorporate less readable financial information into their forecasts. If lower readability increases the costs of processing and interpreting firm disclosures, then it could limit the ability of analysts to correctly incorporate all of the pertinent information and could result in disagreement or ambiguity. We predict and find that analyst earnings forecasts for firms with less readable 10-K reports have greater dispersion, are less accurate, and are associated with greater overall analyst uncertainty. We also find that analyst common uncertainty is increasing in 10-K Fog. These results suggest that analyst forecasts are affected by 10-K readability and provide indirect evidence to support the notion that the information contained in the 10-K filing is used by analysts.

This study contributes to the literature in the following ways. First, our overall findings that lower readability is associated with greater levels of analyst coverage, effort, and information content, but with lower accuracy, higher dispersion, and greater uncertainty of their earnings forecasts complements and contributes to the literature on how analysts respond to firms' disclosure. While prior studies generally focus on the effect of the complexity of specific attributes of firms' disclosure on analyst behavior, we provide evidence that suggests that analyst behavior is associated with overall disclosure readability.[5] While the collection of the evidence in this study is consistent with this inference, the evidence obtained from each individual test is subject to limitations common to our research design.

[5] We note that covering firms with less readable communication is likely to result in a trade-off. Investors are more likely to find analyst coverage useful for firms with less readable communication; however, earnings forecasts for these firms will potentially be less accurate. Our findings suggest that the benefits of coverage outweigh the costs, because we document that both greater coverage and less accurate forecasts are associated with less readable communication (see also Li et al. 2009). This is consistent with the notion that investors value more than just forecast accuracy (Schipper 1991; Clement and Tse 2003).

Second, our findings relate to the debate about the intended audience of public corporate filings. Since its organization, the SEC has made consistent efforts to encourage firms to make their regulatory filings accessible to the average or "lay" investor (Firtel 1999).[6] The most recent of these efforts is the plain English disclosure rules adopted by the SEC on January 22, 1998. Our evidence, that less readable corporate disclosures are associated with greater analyst following and more informative analyst reports, suggests that analysts could serve as an available information intermediary for investors and relates to the SEC's concerns about the accessibility of these reports.

Section II discusses the role of the readability of firm communication and the hypotheses development. Section III describes the data and sample, and section IV presents our empirical evidence. Section V concludes.

II. HYPOTHESES DEVELOPMENT

Less readable communication is more difficult to interpret and process by investors because it requires that investors devote more time and effort to identify and extract relevant information (Bloomfield 2002). In this study we focus on the written complexity of firm disclosures, as measured by readability rather than content. We interpret disclosure readability to be a measure of the costs incurred by users to process and interpret a firm's written communication after controlling for the operational complexity of the business.[7] Firms with similar operations provide disclosures with varying levels of readability. For example, Berkshire Hathaway and AIG both provide a section about their reinsurance businesses in the management's discussion and analysis section of their 10-K filings for fiscal year 2003. While the subject matter and underlying business complexity of the reinsurance business sections are very similar across the two firms, Berkshire Hathaway's explanation is more readable as evidence by a lower Fog Index.[8] This is not surprising, insofar as Berkshire Hathaway CEO Warren Buffett is a strong proponent of more readable communication.

We begin our analysis by examining the effect of disclosure readability on analyst following. We assume that users of financial information have different abilities to process complex communication (Indjejikian 1991; Ball 1992). These differences provide opportunities for information intermediaries, such as financial analysts, to profit from their private analysis of firms by selling their opinions to users with greater information-processing costs (Schipper 1991). These profitable opportunities are arguably greater for firms with less readable disclosures because of the greater cost to users of processing the firms' information. As such, if analysts respond to this increased demand for their services, then we expect analyst following to be greater for firms with less readable disclosures.

---

6 Some academic scholars and practitioners have argued that the primary users of this information should be market professionals such as analysts and not the "lay" investor (Kripke 1970; Schipper 1991). This is because nonprofessionals might not possess the skill or expertise to read and understand the complex financial information contained in disclosure documents and, therefore, any effort to gear disclosure toward the layperson is a waste of time and money (Kripke 1970). Further, critics argue that the SEC's efforts to appeal to the average investor are not only inefficient, but also are hazardous to the disclosure regime because simplified and concise disclosure often leaves out many issues that are potentially valuable to the professional and leads to potential legal liability.

7 It is important to note that more readable communication can come at a cost. Legal scholars point out that "it is much harder to simplify than to complicate" (Kimble 1994, 53) and it requires significant skill, work, and time to compose documents in plain language. Additionally, some critics argue that plain English communication increases the risk of litigation (Kripke 1973).

8 The Fog Index for AIG's reinsurance business section of the MD&A is 18.51, while Berkshire Hathaway's is 17.23. The difference between these scores (1.28) can be interpreted as the number of additional years of formal education required to understand the text on a first reading. Further details regarding the Fog Index are provided in Section III.

We note, however, that analysts also face a variety of costs in covering firms with less readable communication. First, analysts bear the direct costs of processing the information provided by management. Second, less readable communication can result in confusion, requiring analysts to incur greater private search costs as they obtain additional information to evaluate and interpret management's communication. Third, greater syntactic complexity could lead to inaccurate forecasts and recommendations that adversely affect analysts' careers (e.g., Mikhail et al. 1999; Plumlee 2003; Hong and Kubik 2003).[9] Finally, Li (2008) finds that firms attempt to obfuscate bad news by increasing the complexity of their communication, and Lang et al. (2004) find that analysts are less likely to cover firms with incentives to withhold or manipulate information. These potential manipulations along with the costs to analysts of covering firms with less readable communication discourage analyst following. Because there are both potential positive and negative consequences of syntactic complexity on analyst following, we test the null hypothesis that syntactic complexity has no association with analyst following.

Given the difficulty of following firms with less readable disclosures, analysts who choose to follow these firms likely exert greater effort to do so. One way of measuring the effort that analysts exert in following firms with less readable disclosures is to examine the average time from the firm's 10-K filing to the analyst's first report subsequent to the filing. If lower readability results in higher processing costs to analysts, then firms with less readable communication will have greater average response time from the analysts that cover them as compared to the analysts covering firms with more readable communication. Similarly, analysts who cover a portfolio of firms that have less readable communication should take longer, on average, to issue their reports than other analysts. We test the hypothesis that analysts exert greater individual effort to cover firms with less readable disclosures.

Readability could also affect the properties of analyst earnings forecasts. Specifically, we examine its effect on the information content, accuracy, and dispersion of analyst forecasts and the uncertainty in analysts' information environment. Analyst reports for firms with less readable communication likely provide information that is more useful to investors due to greater costs in processing public information. Analyst reports for firms with less readable communication would also be more informative if analysts choose to acquire and incorporate more private information due to the difficulties involved in interpreting less readable disclosures. Accordingly, we predict that analyst reports for firms with less readable communication will have greater information content, on average, than firms with more readable communication.

Finally, because analysts who follow firms with less readable communication bear the costs of processing and interpreting such disclosures, syntactic complexity could also affect the dispersion and accuracy of analyst earnings forecasts as well as the degree of the overall and common uncertainty embedded in analyst forecasts. If less readable communication increases analysts' cost to process and interpret the information, then it is likely to lead to a more diverse set of interpretations about firm disclosures, resulting in higher analyst forecast dispersion. In addition, if less readable communication makes it more difficult to forecast earnings, then the accuracy of the analyst consensus forecast will be lower for firms with less readable communication. However, this difference in accuracy can be fully or partially offset by greater analyst effort in response to greater syntactic complexity. Accordingly, we predict that analyst forecasts will have greater dispersion and be less accurate for firms with less readable communication.

---

[9] Li et al. (2009) find that once analysts become all-stars they are more willing to cover firms with greater potential earnings management because investors could find their coverage more valuable for these firms. This suggests that coverage decisions can be related to favorable career outcomes; however, it is not clear how analysts assess the trade-offs between the demand for investment information and the cost of potentially less accurate forecasts.

Barron et al. (1998) point out that analyst earnings forecast dispersion and accuracy do not directly capture the theoretical properties of analysts' information environment such as uncertainty because these measures relate in different ways to the idiosyncratic and the common components of error in analysts' forecasts. They view idiosyncratic uncertainty as uncertainty in private information that affects each analyst differently, and common uncertainty as uncertainty in information that is common to all analysts. Accordingly, lower readability can be associated with higher dispersion and lower accuracy in earnings forecasts for reasons related to either or both common and idiosyncratic uncertainty in the analysts' information environment.

To further examine the association between readability and properties of analyst earnings forecasts, we follow the model derived in Barron et al. (1998) and construct empirical measures of the overall uncertainty (modeled as the sum of the idiosyncratic and the common uncertainty) and the common uncertainty in analyst forecasts. These measures combine the accuracy, the dispersion, and the number of analyst forecasts, allowing us to examine more directly how readability relates to analysts' information environment uncertainty. We predict that analyst forecasts for firms with less readable reports will be associated with greater overall uncertainty. The effect of readability on common uncertainty (i.e., how much the average belief reflects common versus private information) is less clear because analysts can make trade-offs between public and private information sources based on the relative precision and importance of these sources.

III. SAMPLE AND VARIABLE DEFINITIONS

Sample Selection

Our initial sample is based on the intersection of firm/years available on the Compustat Fundamental Annual table and the SEC's EDGAR filings database for fiscal years 1995–2006. These databases are joined based on Compustat GVKEY and the SEC's Central Index Key (CIK). Firms without matches are dropped from the sample. For each firm-year observation, we download the corresponding 10-K filing. Filings with less than 3,000 words or 100 lines are dropped to ensure that a complete filing is examined and that no errors were made in the filing transmission. This procedure results in 57,642 observations. We obtain stock return data from CRSP, analyst data from I/B/E/S, institutional holdings data from Thomson Reuters (CDA/Spectrum), and information on management earnings guidance from the First Call Company Issued Guidelines database. This procedure yields a sample of 33,704 observations.

Disclosure Readability

Similar to Li (2008), we measure the readability of 10-K filings using the Fog Index (*FOG*). This index, developed in the computational linguistics literature, captures the written complexity of a document as a function of the number of syllables per word and the number of words per sentence. Specifically, we calculate the readability of the 10-K report for firm i in year t as follows:

$$FOG_{i,t} = (\text{average words per sentence} + \text{percent of complex words}) \times 0.4 \tag{1}$$

where a complex word is defined as one with three or more syllables. The index is interpreted as the number of years of formal education required for a person of average intelligence to read the document once and understand it.[10] It is important to note that the Fog Index is a measuring tool,

[10] Before computing the Fog Index, we remove all tables, tabulated text, and financial statements from the 10-K. The constant of 0.4 found in Equation (1) was chosen by Robert Gunning based on the scores of a set of literary benchmarks in order that this specific interpretation could be made. Our inferences remain unchanged using alternative measures of readability such as the Kincaide grade level formula, the Flesch Reading Ease Index, and 10-K length. In addition, our main inferences are similar after including in the analysis the length of the 10-K filing as a control variable for the amount of disclosure provided by the firm (Leuz and Schrand 2009).

not a rule or formula for good writing; as stated by Gunning (1969, 12), "Nonsense written simply is still nonsense." It predicts the readability of a document, but does not provide information about whether the writing is interesting or informative. Despite these limitations, it is objective and simple to calculate. It allows us to study the disclosure characteristics of a large and diverse group of firms and does not depend on analyst surveys or opinions. It also provides us with a comprehensive measure of the overall syntactic complexity of 10-K filings as opposed to the complexity of individual financial items.[11]

Table 1 provides descriptive statistics on the Fog Index for firms in the intersection of Compustat and EDGAR. As can be seen in Panel A, the overall mean and median of the Fog Index are 19.52 and 19.38, respectively. As a comparison to help validate our measure, it is interesting to note that the average Fog Index for FASB Statement Nos. 1–122 is 22, the CPA Exams is 16, the CMA Exams is 17, *Wall Street Journal* is 12, and *Reader's Digest* is 8 (Phillips et al. 2007; Cox 2007). The relatively high mean and median Fog Index in our sample is consistent with concerns that financial reports are written in complex language.

While the variation in the mean Fog Index over the sample period is modest, there is large variation within each year. For example, over the total sample years, the interquartile range for the Fog Index is from 1.58 to 1.88 (about 9 percent of the mean). There is also significant variation in the Fog Index within industries, despite similarities in the underlying business complexity within each industry. Table 1, Panel B provides examples of specific industries that have high and low levels of the Fog Index (industries are classified using the Fama and French 48-industry classification). The healthcare, insurance, trading (e.g., security brokers, investment offices, etc.), utilities, and telecommunications industries comprise the group with the highest Fog Index. The fact that these industries have a low level of disclosure readability is not surprising because they are characterized by complex contracts and business models that are difficult to communicate. However, there is also significant variation within each industry. Industries with a low Fog Index comprise precious metals, shipping containers, food products, agriculture, and defense industries. While the precious metals and shipping container industries have small within-industry variation, most industries have an interquartile range of the Fog Index above 1.60. The relative rankings of these industries also help to validate the ability of the Fog Index to measure the readability of annual reports.

Variable Definitions

Our tests focus on examining the relation between the Fog Index and analyst following, their forecast revision response time, the information content of their reports, and the properties of their earnings forecasts. A description of the construction of these variables as well as our control variables follows.

Analyst Following

Similar to prior research (O'Brien and Bhushan 1990; Brennan and Subrahmanyam 1995), we define analyst following as the number of analysts (*#ANALYSTS*) that comprise the first I/B/E/S consensus annual earnings forecast after the filing date of the 10-K report. We follow Bhushan (1989) and interpret this measure as a proxy for the collective effort of the financial analyst community in the analysis of an individual firm.[12] Because some firms are not covered by I/B/E/S,

[11] Prior literature used the Fog Index and other similarly constructed measures to examine the readability of the overall annual report (Jones and Shoemaker 1994), management's discussion and analysis (Schroeder and Gibson 1990), and the notes to the financial statements (Smith and Smith 1971; Healy 1977).

[12] Similar to Bhushan (1989), we acknowledge that our proxy is not a perfect measure because it assumes homogeneity among analyst effort levels. For example, there are differences in individual analysts' effort levels based on differences in compensation, brokerage houses, etc.

TABLE 1

Descriptive Statistics for Disclosure Readability Using the Fog Index of the 10-K Filing

Panel A: Fog Index by Fiscal Year (All Compustat Firms)

Year	n	Mean	Std. Dev.	Q1	Median	Q3	Interquartile
1995	3,107	19.25	1.31	18.38	19.17	20.04	1.66
1996	5,221	19.45	1.31	18.53	19.38	20.24	1.71
1997	5,530	19.48	1.30	18.59	19.39	20.28	1.70
1998	5,329	19.44	1.25	18.59	19.35	20.20	1.62
1999	5,524	19.26	1.29	18.40	19.14	20.02	1.62
2000	5,494	19.19	1.29	18.34	19.05	19.92	1.58
2001	4,955	19.30	1.42	18.39	19.14	20.00	1.61
2002	4,723	19.57	1.52	18.60	19.38	20.33	1.73
2003	4,546	19.91	1.79	18.80	19.63	20.68	1.88
2004	4,425	19.78	1.86	18.81	19.64	20.67	1.87
2005	4,449	19.86	1.50	18.88	19.68	20.56	1.68
2006	4,339	19.90	1.49	18.94	19.72	20.61	1.67
All Years	57,642	19.52	1.47	18.58	19.38	20.29	1.70

Panel B: Fog Index by Industry (All Compustat Firms)

	n	Mean	Std. Dev.	Q1	Median	Q3	Interquartile
Industries with High Fog Index							
Healthcare	1,115	20.22	1.46	19.30	20.04	20.99	1.69
Insurance	2,143	20.16	1.29	19.32	20.10	20.89	1.57
Trading	3,279	19.88	1.67	18.78	19.75	20.75	1.97
Utilities	2,405	19.84	1.56	18.92	19.64	20.53	1.61
Telecommunications	2,021	19.80	1.45	18.89	19.59	20.45	1.56
Industries with Low Fog Index							
Precious Metals	135	18.43	0.94	17.75	18.38	19.12	1.37
Shipping Containers	159	18.64	1.25	17.87	18.43	19.20	1.33
Food Products	483	18.86	1.56	17.95	18.65	19.63	1.68
Agriculture	162	18.95	1.47	17.90	18.85	19.58	1.68
Defense	108	18.99	1.45	18.00	18.87	19.88	1.88

(continued on next page)

Panel A reports descriptive statistics on our measure of disclosure readability, the Fog Index, for observations available on the SEC's EDGAR and Compustat databases. The Fog Index is computed as: (average words per sentence + percent of complex words) $\times$ 0.4, using the text of the 10-K filings for fiscal years 1995–2006.

Panel B provides descriptive statistics for the five industries with the highest and lowest mean Fog Index. Industries are based on the Fama and French 48-industry classification.

we conduct tests for both the I/B/E/S sample and the full sample of firms, where missing coverage is coded as zero analyst coverage (Barth et al. 2001). Because these results are similar, we report our main results based on the I/B/E/S sample.

Analyst Forecast Revision Response Time

We define the analyst forecast revision response time as the time from the 10-K filing to the first annual or quarterly earnings forecast issued by each *individual* analyst following the firm. To ensure that we include only analysts who actively follow the firm, we require that each analyst issue a forecast in the 90 days prior to the 10-K filing and then another report within 90 days of the 10-K filing (this process eliminates 10,458 observations related to analysts who stopped coverage). Because earnings announcements can prompt analysts to issue reports, we exclude reports made after any earnings announcement that occurs after the 10-K filing, but before the end of the 90-day window. We then define analyst report duration as the length of time in days between the 10-K filing and the first report following the filing. In addition, we average individual analyst report duration at both the individual firm and analyst level for each year. We interpret this duration variable as a measure of the required amount of effort for an individual analyst to read, understand, and process the information contained in the 10-K filing and to issue an updated earnings forecast. As mentioned above, we expect analyst response time to be longer for less readable 10-K filings.

Information Content of Analyst Reports

Similar to Frankel et al. (2006), we measure the information content of analyst reports as the proportion of a firm's stock returns related to analyst forecast revisions to the total stock return during the time period between the 10-K filing and the subsequent fiscal year-end. This measure is constructed as the sum of the one-day, absolute size-adjusted returns on the analyst forecast revision day, divided by the sum of the one-day, absolute size-adjusted returns over the entire window.[13] We exclude analyst reports that coincide with earnings announcements. We treat multiple reports issued on the same day as a single report. Also, an observation must have a minimum of 90 trading days with available data to be considered in our tests. Similar to Frankel et al. (2006), we interpret this measure as the percentage of total firm information provided to investors that is related to analyst reports.

Properties of Analyst Earnings Forecasts

We define analyst forecast dispersion (*DISPERSION*) as the standard deviation of the individual analyst forecasts in the first analyst consensus annual earnings forecast issued after the 10-K filing for the fiscal period following the 10-K filing, scaled by share price 90 days before the consensus forecast date. Analyst forecast accuracy (*ACCURACY*) is computed as the squared difference between I/B/E/S reported earnings and the analyst consensus forecast, scaled by share price 90 days before the consensus forecast date.[14] We define analyst overall uncertainty

---

[13] Frankel et al. (2006) divide this measure by the number of forecast revision dates to obtain a measure of the average informativeness of an analyst report date. We omit this final step for two reasons. First, our variable of interest is the overall firm information that comes from analysts, not the average information content. Second, firms with less readable disclosures are likely to obtain greater amounts of information from analysts because analysts are more likely to revise their forecasts more often for such firms. We confirm this supposition in untabulated results. Thus, dividing by the number of forecast revisions would induce a negative correlation between the analyst informativeness measure and 10-K readability. Our results are robust to the inclusion of the number of analysts following the firm as a control for the expected number of analyst revisions.

[14] Our inferences are not sensitive to scaling the measures of analyst forecast accuracy and dispersion by the absolute value of earnings or using the unscaled variables.

($UNCERTAINTY_{OVERALL}$) and common uncertainty ($UNCERTAINTY_{COMMON}$) using the following equations derived by Barron et al. (1998):[15]

$$UNCERTAINTY_{OVERALL} = \left(1 - \frac{1}{\#ANALYSTS}\right) * DISPERSION + ACCURACY \tag{2}$$

$$UNCERTAINTY_{COMMON} = \frac{ACCURACY - \dfrac{DISPERSION}{\#ANALYSTS}}{UNCERTAINTY_{OVERALL}}. \tag{3}$$

Following Barron et al. (1998), we interpret overall uncertainty ($UNCERTAINTY_{OVERALL}$) as the sum of the idiosyncratic uncertainty (i.e., uncertainty associated with analysts' private information) and common uncertainty (i.e., uncertainty associated with information common to all analysts). $UNCERTAINTY_{COMMON}$ is calculated as the proportion of common uncertainty to overall uncertainty. Thus, we interpret $UNCERTAINTY_{COMMON}$ to be a measure of how much the average analyst's belief reflects common versus private information. As shown in Equations (2) and (3), as *DISPERSION* approaches 0, $UNCERTAINTY_{OVERALL}$ approaches *ACCURACY* and $UNCERTAINTY_{COMMON}$ approaches 1. This suggests that, when there is no disagreement among analysts (i.e., *DISPERISON* equals 0), the total uncertainty is only associated with analysts' common information and all information impounded in analyst forecasts is public. Empirically, because the uncertainty measures emphasize information across analysts, we require that each firm have at least four analysts following it.

Control Variables

Our analysis controls for a variety of variables that have been shown by prior literature to be associated with firms' information environment and business complexity and therefore relate to analyst behavior. Prior work finds that firm size is the most important determinant of analyst following (Bhushan 1989; O'Brien and Bhushan 1990; Brennan and Hughes 1991; Lang and Lundholm 1996; Barth et al. 2001). These studies find that larger firms have greater analyst following and suggest that large firms have better information environments, potentially more complex operations, and greater demand for investment advice. We use the natural logarithm of market value as of the year ending prior to the 10-K filing (*LOGSIZE*) as a proxy for size.

Following Barth et al. (2001), we include controls for growth. High-growth firms tend to attract greater analyst following due to investor interest and the potential for future investment banking deals. Further, analysts likely find it more difficult to accurately forecast earnings for firms with high growth, leading to greater disagreement among analysts and less accurate forecasts. We define the variable *GROWTH* as the compounded average growth rate in sales over the prior three to five fiscal years. Our inferences are robust to the inclusion of the natural logarithm of the book-to-market ratio as a proxy for growth.

Similar to Bradshaw et al. (2008), we include the natural logarithm of the number of business segments reported in the Compustat Segment File as a control for the underlying complexity of the firm. We also include a control for the level of institutional holdings, following the evidence in Bhushan (1989), Brennan and Subrahmanyam (1995), and Frankel et al. (2006). These studies find that institutional ownership is positively associated with analyst following and with the informa-

[15] We thank the reviewer for suggesting that we incorporate the Barron et al. (1998) measures into our analysis. Note that the variables we term overall and common uncertainty are termed analyst uncertainty and consensus, respectively, in Barron et al. (1998). We use the term common uncertainty as opposed to consensus to avoid confusion regarding our use of analyst consensus forecasts.

tion content of their reports.[16] Institutional ownership can also be associated with higher analyst forecast accuracy and lower dispersion because firms with high levels of institutional holdings tend to have better information environments. We define the variable *PINST* as the percentage of a firm's shares that are held by institutions from the 13F disclosures for the most recent quarter prior to the 10-K filing. Motivated by the evidence in Lang and Lundholm (1996), who document that analyst following increases with the quality of disclosures, we include the number of management earnings forecasts made during the prior year (*MFCOUNT*) as a proxy for firm discretionary disclosure (Nagar et al. 2003; Cotter et al. 2006), as well as the absolute value of the cumulative, market-adjusted return for the two-day window around the 10-K filing (*10-K NEWS*) as a proxy for the disclosure informativeness.[17]

Similar to Barth et al. (2001), we include variables to control for firms' information environment and business complexity, as these variables likely affect the properties of analyst earnings forecasts. Barth et al. (2001) examine the association between analysts' incentives to follow firms and the extent of their intangible assets. Because many intangible assets are generally not recognized and estimates of their fair value are not disclosed, Barth et al. (2001) argue that analysts have increased incentives to follow firms with greater intangibles due to increased demand from investors. They find that analyst following is greater for firms with larger research and development and advertising expenses. They interpret these results as evidence that analysts respond to the demands of investors for more information because of the difficulty in evaluating firms' intangible assets. Other studies examine the effect of intangibles on the properties of analyst forecasts. Barron et al. (2002) find that analyst uncertainty increases with the level of a firm's intangible assets. Gu and Wang (2005) find that analyst forecast errors are increasing in firm intangible intensity. Intangibles are important for our tests because they could be associated with 10-K readability. For example, it is likely more difficult to explain the operations of firms with high levels of research and development costs in a less complex manner. Similar to Barth et al. (2001), we define *R&D* as the ratio of research and development expense to operating expense and *ADV* as the ratio of advertising expense to operating expense.[18]

Finally, we include the standard deviation of firm monthly stock returns from the prior year (*STD_RET*) as a measure of information uncertainty.[19] Bhushan (1989) suggests that private information is more valuable for firms with higher return volatility and thus positively related to the demand for analyst services. However, it is likely that analysts bear increased costs for following firms with higher return volatility. Additionally, most of our tests include industry fixed affects (based on the Fama and French classification) and year fixed effects to account for variation in analyst following across specific industries and over time. This approach is used to help control for variation in business complexity or information uncertainty that is driven by industry or time.

---

[16] O'Brien and Bhushan (1990) suggest that analyst following and institutional ownership are potentially endogenous. Frankel et al. (2006) assume that the variables are exogenous. Similar to Barth et al. (2001), we examine our results with and without the inclusion of institutional ownership and also with the inclusion of lagged institutional ownership. The coefficient on *FOG* remains positive and significant under these specifications.

[17] We thank a reviewer for pointing out the importance of controlling for the informativeness of the 10-K disclosure.

[18] Our results are robust to the inclusion of the amount of recognized intangibles and depreciation expense, which are also examined by Barth et al. (2001).

[19] While some of our other control variables (e.g., sales growth and R&D) are likely correlated with firms' information uncertainty, we include this measure as a more explicit, market-based measure of this construct. In addition, the inclusion of an alternative measure, earnings volatility, does not significantly affect our results. However, the inclusion of this measure reduces our sample by about 20 percent due to the need for a sufficient time-series of earnings.

IV. RESULTS

Summary Statistics

Table 2 presents summary statistics for the sample of 33,704 firm-year observations that remain after imposing the availability of data on CRSP and I/B/E/S. Similar to Table 1 and consistent with concerns raised by the SEC about disclosure complexity, the mean (median) *FOG* score is 19.53 (19.38). The standard deviation and interquartile range are 1.42 and 1.71, respectively, with an interquartile range from 18.58 to 20.29. The mean (median) number of analysts per firm-year observation is 6.14 (4.00). The mean (median) average analyst forecast revision response time is 17.71 (16.00) days at the firm-year level and 18.77 (17.50) days at the analyst-year level. This suggests that, on average, analysts do not respond immediately to 10-K filings due to the significant amount of information that must be processed and interpreted. The mean (median) analyst report information content (*AI*) is 0.130 (0.092). This indicates that, on average, almost 13 percent of the information reflected in stock returns during the period between the 10-K filing and the end of the fiscal period is derived from analyst reports. The mean (median) of forecast dispersion is 0.008 (0.003) and that of squared forecast error (*ACCURACY*) is 0.078 (0.002). The mean (median) of $UNCERTAINTY_{OVERALL}$ and $UNCERTAINTY_{COMMON}$ are 0.051 (0.005) and 0.328 (0.267), respectively. This suggests that about 33 percent of analyst uncertainty about future earnings following the 10-K filing is based on public information.

Table 2 also provides statistics on the control variables. The mean (median) size of our sample firms is $2.1 billion ($381 million), and mean (median) compound averages growth rate of sales (*GROWTH*) is 0.20 (0.11). The mean (median) number of business segment is 1.83 (1), and mean (median) percent of institutional ownership is 46 (45). The mean (median) number of management earnings forecasts is 1.44 (0) and the mean (median) *10-K NEWS* is 0.03 (0.02). The mean (median) ratio of research and development expense (*R&D*) and advertising (*ADV*) to operating expense are 0.08 (0.00) and 0.01 (0.00), respectively.[20] The mean (median) standard deviation of returns (*STD_RET*) is 0.14 (0.12).

Similar to Li (2008), we find in untabulated results that, while *FOG* is significantly correlated with many of the control variables, the extent of these correlations is relatively small for most variables. This is consistent with the conclusion that it is difficult to explain a significant proportion of the variation in *FOG* using firm characteristics. This evidence is important because in this study we assume that the Fog Index measures the readability of 10-K disclosures rather than their content. Consistent with this assertion, the correlation between *FOG* and the absolute value of the 10-K event return is small and insignificant.

As a caveat, we note that disclosure readability can be decomposed into innate and discretionary components. While we include a variety of variables to control for innate readability, analysts' information costs are based on total readability, so the relative magnitude of these components is less relevant. Our inferences are limited by the extent to which the Fog Index measures readability and is not confounded with other (uncontrolled) firm characteristics. It is also important to note that, for a specific variable to influence the readability measure, it must either increase the length of the average sentence in the 10-K document or increase the percentage of complex words. Simply increasing the length of the 10-K will not directly affect the measure. However, we examine the effect of disclosure length on our main results in sensitivity tests because length also can be a form of complexity (You and Zhang 2009; Loughran and McDonald 2010; Miller 2010).

---

[20] Note that the medians for the number of management earnings forecasts, research and development expense, and advertising expense are zero. This is primarily because we follow prior literature and code missing items to be zero.

TABLE 2
Sample Descriptive Statistics

Variable	n	Mean	Std. Dev.	Q1	Median	Q3
FOG	33,704	19.53	1.42	18.58	19.38	20.29
#ANALYSTS	33,704	6.14	5.91	2.00	4.00	8.00
$RESPONSE_{Firm}$	17,868	17.71	10.76	9.50	16.00	24.25
$RESPONSE_{Analyst}$	5,737	18.77	8.49	12.50	17.50	23.50
AI	30,716	0.130	0.139	0.037	0.092	0.177
DISPERSION	26,078	0.008	0.018	0.001	0.003	0.008
ACCURACY	29,055	0.078	0.447	0.000	0.002	0.015
$UNCERTAINTY_{OVERALL}$	17,241	0.051	0.280	0.002	0.005	0.016
$UNCERTAINTY_{COMMON}$	17,222	0.328	0.401	−0.038	0.267	0.702
SIZE	33,704	2100	5600	122	381	1300
GROWTH	33,704	0.20	0.34	0.04	0.11	0.24
SEGMENTS	33,704	1.83	1.32	1.00	1.00	3.00
PINST	33,704	45.98	26.78	22.95	45.31	67.60
MFCOUNT	33,704	1.44	2.69	0.00	0.00	2.00
10-K NEWS	33,704	0.03	0.05	0.01	0.02	0.04
ADV	33,704	0.01	0.03	0.00	0.00	0.01
R&D	33,704	0.08	0.17	0.00	0.00	0.08
STD_RET	33,704	0.14	0.10	0.07	0.12	0.18

This table reports descriptive statistics for observations that are available on the SEC's EDGAR, Compustat, CRSP, and I/B/E/S databases. The variables are pooled across fiscal years 1995–2006.

Variable Definitions:

FOG = Fog Index of the 10-K filing calculated as (average words per sentence + percent of complex words) × 0.4;
#ANALYSTS = number of analysts in the first consensus annual earnings forecast following the 10-K filing;
$RESPONSE_{Firm}$ = average number of days that it takes a firm's analysts to issue their first report following the 10-K filing;
$RESPONSE_{Analyst}$ = average number of days that it takes an analyst to issue his/her first report following the 10-K filing;
AI = sum of the one-day, absolute size-adjusted returns from analyst reports between the 10-K filing and the next fiscal year-end divided by the sum of the one-day, absolute size-adjusted returns over the entire window;
DISPERSION = standard deviation of the individual analyst forecasts in the first analyst consensus annual earnings forecast issued after the 10-K filing for the fiscal period following the 10-K filing, scaled by share price 90 days before the consensus forecast date;
ACCURACY = squared difference between I/B/E/S reported earnings and the first analyst consensus annual earnings forecast issued after the 10-K filing for the fiscal period following the 10-K filing, scaled by share price 90 days before the consensus forecast date;
$UNCERTAINTY_{OVERALL}$ = sum of common and idiosyncratic uncertainty among analysts computed following Barron et al. (1998, equation 15), using the measures of accuracy, dispersion, and analyst following as previously defined in this table;
$UNCERTAINTY_{COMMON}$ = ratio of common uncertainty to total uncertainty among analysts computed following Barron et al. (1998, equation 16) using the measures of accuracy, dispersion, and analyst following as previously defined in this table;
SIZE = number of shares outstanding (Compustat item CSHO) times the share prices at the most recent fiscal year-end (Compustat item PRCC_F);
GROWTH = compound average growth rate of firm sales (Compustat item SALE) over the prior 3–5 years $(sales_{t-1} / sales_{t-(1+i)})^{1/i}$;
SEGMENTS = number of reported business segments in the Compustat segment file for prior fiscal year;
PINST = percentage of institutional ownership from the quarter prior to the 10-K filing;

(continued on next page)

TABLE 2 (continued)

MFCOUNT = number of management earnings forecasts issued in the prior year;
10-K NEWS = absolute value of the cumulative market-adjusted return for the 10-K filing event window [0,1];
ADV = advertising expense (Compustat item XAD) as a percentage of operating expense (Compustat item XOPR) from the prior fiscal year;
R&D = research and development expense (Compustat item XRD) as a percentage of operating expense (Compustat item XOPR) from the prior fiscal year; and
STD_RET = standard deviation of the firm's monthly stock returns from the previous fiscal year.

Analyst Following and the Readability of 10-K Filings

Our first prediction is that analyst following is affected by the level of readability of firms' 10-K filings, as measured by the Fog Index. To control for other factors that can affect analyst following, we estimate the following regression:

$$\begin{aligned}\#ANALYSTS_{i,t} = {} & \beta_0 + \beta_1 FOG_{i,t} + \beta_2 LOGSIZE_{i,t} + \beta_3 GROWTH_{i,t} + \beta_4 LOGSEGMENTS_{i,t} \\ & + \beta_5 PINST_{i,t} + \beta_6 MFCOUNT_{i,t} + \beta_7 10\text{-}K\ NEWS_{i,t} + \beta_8 ADV_{i,t} + \beta_9 R\&D_{i,t} \\ & + \beta_{10} STD\_RET_{i,t} + \varepsilon_{i,t}. \qquad (4)\end{aligned}$$

The estimation is performed using ordinary least-squares regression with industry and time fixed effects.

Column 1 of Table 3 (model 1) reports the results of the linear model regression. t-statistics, presented in brackets, are based on standards errors that are robust to heteroscedasticity and are clustered at the firm level. The coefficient on *FOG* is positive and statistically significant, suggesting that analyst following is greater for firms with less readable disclosures. This is consistent with the notion that analysts respond to investors' demand for investment information for firms whose disclosures are more costly to process. The coefficients on the control variables are consistent with prior research. Larger firms are associated with greater analyst following as well as firms with higher growth rates and higher institutional ownership. Consistent with Lang and Lundholm (1996), we find that disclosure practice (*MFCOUNT*) and disclosure informativeness (*10-K NEWS*), are positively related to analyst following. We find that analyst following is negatively associated with the number of business segments.[21] Similar to Barth et al. (2001), we document that analyst following is greater for firms with higher amounts of advertising and research and development expenses. Consistent with Bhushan (1989), we also find that analyst following is positively associated with firm stock return volatility, suggesting that analysts provide greater support to investors when private information is valuable.

In addition to examining the linear association between *FOG* and analyst following, we also examine the association of analyst following and *FOG* using nonlinear and semi-parametric specifications (second and third columns of Table 3, respectively). In the nonlinear specification (model 2), we add a square term (FOG^2) to Equation (2) to capture the possibility that the association between readability and analyst following differs based on the level of *FOG*. In model 3, we examine a semi-parametric specification by including indicator variables for the various quartiles of *FOG*. The coefficient on FOG^2 (model 2) is negative and significant, indicating that the role of

[21] This effect becomes negative after controlling for firm size, but is positive if size is omitted. The negative coefficient in the multivariate results is consistent with the findings of Bhushan (1989) and Johnston et al. (2009).

TABLE 3

The Association between Annual Report Readability and Analyst Following

Panel A: OLS Regression

Variable	Model 1	Model 2	Model 3
Intercept	−13.067***	−20.190***	−11.625***
	[−16.79]	[−6.29]	[−18.23]
FOG	0.085***	0.798***	
	[3.74]	[2.58]	
FOG^2		−0.018**	
		[−2.33]	
FOG Q2			0.168**
			[2.41]
FOG Q3			0.331***
			[4.11]
FOG Q4			0.369***
			[4.18]
LOGSIZE	2.623***	2.624***	2.623***
	[62.20]	[62.23]	[62.22]
GROWTH	0.361***	0.356***	0.356***
	[3.90]	[3.85]	[3.85]
LOGSEGMENTS	−0.367***	−0.368***	−0.368***
	[−4.68]	[−4.70]	[−4.70]
PINST	0.011***	0.011***	0.011***
	[5.20]	[5.20]	[5.22]
MFCOUNT	0.086***	0.086***	0.086***
	[5.82]	[5.82]	[5.83]
10-K NEWS	2.391***	2.380***	2.371***
	[5.46]	[5.43]	[5.40]
ADV	7.555***	7.564***	7.553***
	[4.57]	[4.57]	[4.57]
R&D	1.696***	1.673***	1.662***
	[5.61]	[5.54]	[5.50]
STD_RET	0.932***	0.924***	0.931***
	[2.61]	[2.59]	[2.61]
Time Fixed Effects	Yes	Yes	Yes
Industry Fixed Effects	Yes	Yes	Yes
n	33704	33704	33704
Adjusted R^2	0.63	0.63	0.63

Panel B: Comparative Interquartile Effects

	Model 1	Model 2	Model 3
FOG	0.146	0.188	0.201
LOGSIZE	6.236	6.238	6.238
GROWTH	0.071	0.070	0.070
LOGSEGMENTS	−0.403	−0.404	−0.405
PINST	0.473	0.473	0.478
MFCOUNT	0.172	0.172	0.172
10-K NEWS	0.081	0.080	0.080
ADV	0.042	0.042	0.042

(continued on next page)

Panel B: Comparative Interquartile Effects

	Model 1	Model 2	Model 3
RD	0.130	0.129	0.128
STD_RET	0.098	0.097	0.174

**, *** $p < 0.05$ and $p < 0.01$, respectively, two-tailed t-tests.

Panel A reports coefficient estimates and t-statistics (in brackets) from the regression of analyst following on *FOG* and control variables. Analyst following is defined as the number of analysts contained in the first I/B/E/S consensus annual earnings forecast following the 10-K filing. *FOG QX* is an indicator variable set to 1 if the firm's 10-K Fog Index is in quartile *X* of the sample, and 0 otherwise. All other variables are as defined in Table 2. Industry fixed effects are based on the Fama and French 48-industry classification. t-statistics are robust to heteroscedasticity and clustered at the firm level. Panel B reports the estimated effect of an interquartile change in the variable of interest.

readability increases at a declining rate. Also, the coefficients on the second, third, and fourth *FOG* quartiles (model 3) are positive and significant and monotonically decreasing across the quartiles. These results are consistent with analysts making a trade-off between the benefits of covering firms with less readable disclosures and the costs of following them. While these costs include the processing costs of covering firms with less readable disclosures, analysts likely find it more difficult to accurately forecast the earnings of firms with more complex disclosures. These costs can be important, as prior research has found that forecast accuracy has important career effects for analysts (Mikhail et al. 1999; Hong and Kubik 2003). The effects of *FOG* on the properties of analyst earnings forecasts will be examined later.[22]

Table 3, Panel B compares the effects of an interquartile change in the independent variables on analyst following based on the marginal effect estimates in Panel A. Consistent with prior literature, size (*LOGSIZE*) has the largest association with analyst following. An interquartile change in the natural logarithm of size is associated with an increase of 6.2 analysts, while that of all other variables is less than 1. The interquartile effect of *FOG* is 0.15 in model 1, 0.19 in model 2, and 0.20 in model 3. While the incremental effect of the complexity of 10-K filings on analyst following is moderate, it is greater than that of most of the other variables included in the regressions. These are the same variables that have been shown in the literature to be related to analyst following. Based on these comparisons, we conclude that disclosure readability is an important determinant of analyst following.

Similar to Lang and Lundholm (1996), we also examine the association between *changes* in *FOG* and *lead changes* in analyst following. If analysts do, in fact, respond to changes in *FOG* by increasing their coverage, then we expect to observe a positive relation between current changes in *FOG* and future changes in analysts following. We investigate this relation by estimating multivariate models using the changes in the independent variables from Table 3. We define the current change in the Fog Index and other independent variables as the difference between their current values and those of the prior fiscal year. We define the lead change in analyst following as the difference between the number of analysts following the firm after the next 10-K filing and the number of analysts following the firm after the current report. Table 4 presents results from these tests. We find that contemporaneous changes in *FOG* are positively related to lead changes in analyst following. We also find that changes in firm size and institutional ownership are positively

[22] In a related study, You and Zhang (2009) examine stock returns following 10-K filings and find that investors underreact to firms with more complex annual filings. This apparent underreaction provides additional motivation for analyst coverage because prior work has found that analysts help improve market efficiency (Barron et al. 2002; Brown and Sivakumar 2003; Gu and Chen 2004).

TABLE 4

The Association between Changes in Annual Report Readability and Lead Changes in Analyst Following

Variable	Model 1	Model 2	Model 3
Intercept	−0.038***	−0.051***	−0.073***
	[−3.79]	[−4.68]	[−3.03]
ΔFOG	0.024***	0.023**	
	[2.66]	[2.55]	
$(\Delta FOG)^2$		0.007**	
		[2.58]	
ΔFOG $Q2$			−0.009
			[−0.28]
ΔFOG $Q3$			0.045
			[1.36]
ΔFOG $Q4$			0.110***
			[3.20]
$\Delta LOGSIZE$	0.480***	0.479***	0.480***
	[21.89]	[21.83]	[21.91]
$\Delta GROWTH$	0.048	0.045	0.050
	[0.66]	[0.62]	[0.68]
$\Delta LOGSEGMENTS$	−0.103**	−0.103**	−0.102**
	[−2.12]	[−2.11]	[−2.11]
$\Delta PINST$	0.008***	0.008***	0.008***
	[5.89]	[5.82]	[5.84]
$\Delta MFCOUNT$	−0.015	−0.015	−0.014
	[−1.61]	[−1.61]	[−1.59]
Δ*10-K NEWS*	−0.138	−0.138	−0.136
	[−0.91]	[−0.91]	[−0.90]
ΔADV	0.333	0.327	0.323
	[0.50]	[0.49]	[0.48]
$\Delta R\&D$	−0.075	−0.075	−0.074
	[−0.57]	[−0.57]	[−0.56]
ASTD_RET	−0.637***	−0.631***	−0.631***
	[−5.18]	[−5.14]	[−5.14]
n	27226	27226	27226
Adjusted R^2	0.03	0.03	0.03

**, *** $p < 0.05$ and $p < 0.01$, respectively, two-tailed t-tests.

This table reports coefficient estimates and t-statistics (in brackets) from the regression of lead changes in analyst following on changes in *FOG* and control variables. Lead change in analyst following is the difference between the number of analysts following the firm after the next 10-K filing and the number of analysts following the firm after the current report. ΔFOG QX is an indicator variable set to 1 if the current change in a firm's 10-K Fog Index is in quartile X of the sample, and 0 otherwise. All other variables are current period changes of the variables defined in Table 2. t-statistics are robust to heteroscedasticity and clustered at the firm level.

related to lead changes in analyst following and that changes in management earnings guidance, number of business segments, and return volatility are negatively related.[23] In untabulated results, we find that *FOG* has the third-highest effect of the variables employed in our analysis. The results from this test are important because they help to alleviate concerns about endogeneity or prior period information shocks relating to the levels results in Table 3.

Analyst Report Duration

One potential measure of the costs or effort that analysts bear in following firms with less readable disclosures is the amount of time it takes them to issue reports following the 10-K filing. We label this length of time "analyst report duration." In this section, we examine the analyst report duration at both the analyst and firm level. We expect that analyst report duration will be longer for firms with less readable 10-K filings. As noted previously, to ensure that we capture analyst reports issued in response to the filing of the 10-K (and not in response to other corporate events), we include in this analysis only firm-analyst observations for which the analyst has issued at least one report during the 90 days prior to the 10-K filing and issues a report in the 90 days subsequent to the 10-K filing.[24] Table 5, Panel A presents univariate results based on individual analyst reports and the Fog Index of the firms they cover. We classify each firm-specific *FOG* as high (low) if it is greater (less) than the median *FOG* value for all firms in the sample. We find that analysts covering firms with a high *FOG* require 1.45 days longer, on average, to issue their reports. Panels B and C of Table 5 present the average analyst report duration per firm-year and analyst-year, respectively. The value of *FOG* at the firm level is the individual firm 10-K Fog for each specific year. The *FOG* score at the analyst level is the average 10-K Fog score of the firms that the analyst covered in each year.[25] The univariate results suggest that the average (median) analyst report duration for firms with *FOG* scores higher than the median is 1.11 (1.50) days (Panel B). At the analyst level, we find that analysts who cover a portfolio of firms with an average *FOG* score higher than the median analyst portfolio average take 2.29 (2.17) days longer on average (median) to issue their first reports. These differences are statistically significant at a 1 percent level and are likely to be economically important given the rapid pace at which markets impound new information.

To further examine the analyst report duration, we estimate ordinary least-squares regressions on the average analyst report durations at both the firm and analyst level. Table 5, Panel D presents the results of these analyses. Similar to the univariate results, we find that firms with higher Fog scores and analysts who cover stocks with higher Fog scores are associated with longer analyst report duration. Specifically, controlling for a variety of other factors that can affect firms' information environment, the coefficient on *FOG* is 0.345 for the firm-level analysis and 1.138 for the analyst-level analysis, both of which are statistically significant. This suggests that firms and analyst portfolios with one unit higher of *FOG* have analyst response times that are 0.345 and 1.138 days longer. Although it is difficult to assess the economic significance of these results, we note that given the pace with which new analyst information is incorporated in stock prices, even such short delays in producing a report could reduce the information content of the report. The results in Table 5 further indicate that firms with greater institutional ownership, discretionary

[23] While less intuitive, one potential concern is that analyst following might lead *FOG*. For example, management might provide more complex information in response to a greater analyst following. In untabulated tests we find no relation between changes in *FOG* and lag analyst following.

[24] As noted previously, we exclude analyst observations made after earnings announcements that follow the 10-K filing and reports made after 90 days after the 10-K. Our results are similar if we use reports within a year after the 10-K filing; however, these observations are likely unrelated to the *FOG* score of the 10-K filing.

[25] To be included in our analysis each analyst must cover a minimum of four firms. Similar results are obtained by requiring only a single firm.

TABLE 5
The Association between Annual Report Readability and Analyst Report Duration

Panel A: Individual Analyst Reports

Group	n	Mean	Median
High Fog Index	30,892	19.19	16.00
Low Fog Index	30,884	17.74	15.00
Difference		1.45***	1.00***

Panel B: Average Analyst Report Duration Per Firm

Group	n	Mean	Median
High Fog Index	8,701	18.33	16.75
Low Fog Index	8,701	17.22	15.25
Difference		1.11***	1.50***

Panel C: Average Analyst Report Duration Per Analyst

Group	n	Mean	Median
High Fog Index	2,869	19.91	18.67
Low Fog Index	2,868	17.62	16.50
Difference		2.29***	2.17***

Panel D: Regression Analysis

	Mean Duration at Firm Level	Mean Duration at Analyst Level
Intercept	9.430†††	−8.750†††
	[7.62]	[−2.62]
FOG	0.345†††	1.138†††
	[5.79]	[6.96]
LOGSIZE	0.034	−0.013
	[0.51]	[−0.11]
GROWTH	−0.453	2.634†††
	[−1.62]	[3.93]
LOGSEGMENTS	0.405††	0.844†††
	[2.57]	[2.67]
PINST	0.039†††	0.105†††
	[9.22]	[11.50]
MFCOUNT	0.170†††	0.019
	[5.68]	[0.33]
10-K NEWS	−25.409†††	−43.743†††
	[−12.33]	[−8.01]
ADV	10.053††	30.936†††
	[2.52]	[4.06]
R&D	1.925†††	0.544
	[3.06]	[0.56]
STD_RET	−6.300†††	−13.345†††
	[−5.53]	[−5.86]
n	17868	5737
Adjusted R^2	0.04	0.11

(continued on next page)

*** $p < 0.01$, two-tailed t-tests for differences of means and Wilcoxon for medians.

††† $p < 0.01$, two-tailed t-tests.

Analyst report duration is defined as the length of time between the 10-K filing and the first report for each analyst. Panels A, B, and C report descriptive statistics of the differences in the average analyst report duration at the individual analyst report, firm-year, and analyst-year levels. The firm and analyst levels are computed by averaging variables over the unit of observation. The High and Low groups are based on the median of the sample in each group. Statistical differences of means (medians) are computed based on t-tests (Wilcoxon tests).

Panel D reports coefficient estimates based on ordinary least square regressions at both the firm and analyst levels. All variables are defined as in Table 2. On the analyst level, the variables are averaged over the portfolio of each analyst-year observation. t-statistics (in brackets) are robust to heteroscedasticity and clustered at the firm and analyst levels, respectively.

disclosure, and firm intangibles are associated with longer average analyst response times. In contrast, firms with greater reaction to the release of the 10-K report and greater prior return volatility are associated with shorter analyst report response time.[26] Overall, the evidence is consistent with the notion that analysts exert more effort to cover firms with less readable disclosures.

Information Content of Analyst Reports

Our third hypothesis predicts that the informativeness of analyst reports is positively related to the readability of firms' 10-K reports. To test this hypothesis, we estimate the following ordinary least-squares regression with industry and time fixed effects:

$$AI_{i,t} = \beta_0 + \beta_1 FOG_{i,t} + \beta_2 LOGSIZE_{i,t} + \beta_3 GROWTH_{i,t} + \beta_4 LOGSEGMENTS_{i,t} + \beta_5 PINST_{i,t} + \beta_6 MFCOUNT_{i,t} + \beta_7 10\text{-}K\ NEWS_{i,t} + \beta_8 ADV_{i,t} + \beta_9 R\&D_{i,t} + \beta_{10} STD\_RET_{i,t} + \varepsilon_{i,t}. \quad (5)$$

Similar to the regression reported in Table 3, we also examine the association of the information content of analyst reports and *FOG* using nonlinear and semi-parametric specifications.

Table 6 reports the regression findings. t-statistics, presented in brackets, are based on standard errors that are robust to heteroscedasticity and clustered at the firm level. Consistent with our hypothesis, the coefficient on *FOG* is positive and significant, suggesting that the informativeness of analyst reports is increasing in the complexity of the 10-K disclosure. This evidence is consistent with the notion that investors find analyst reports for firms with less readable disclosures more useful because of greater processing costs. Analysts' private information searches could also be more valuable in such cases. In model 2, the square term (FOG^2) is negative but insignificant, whereas the coefficients on the individual *FOG* quartiles in model 3 are significant and monotonically increasing. These findings are consistent with our prior results and support the notion that analysts trade off costs and benefits in their coverage of firms with less readable disclosure. The effects of the control variables on the information content of analyst reports are similar in direction to their effects on analyst following. Specifically, the informativeness of analyst reports is increasing in firm size, growth, institutional ownership, discretionary disclosure, firm intangibles, and return volatility. In untabulated results, we find that an interquartile change in *FOG* increases the

[26] In untabulated results, we include a control for the number of reports issued by an analyst for the analyst level results. The coefficient on this variable is not statistically significant and does not quantitatively affect the results as presented. We also estimate two proportional hazard models to examine the influence of readability on the time to an analyst's first report by conditioning at the firm and analyst level, similar to O'Brien et al. (2005). We find that, on average, the probability of an analyst making a report at any given time is between 1.7 percent and 3.2 percent lower for each unit change of *FOG*.

TABLE 6

The Association between Annual Report Readability and the Information Content of Analysts' Reports

Variable	Model 1	Model 2	Model 3
Intercept	−0.256***	−0.302***	−0.227***
	[−16.90]	[−4.09]	[−20.43]
FOG	0.002***	0.006	
	[3.25]	[0.86]	
FOG^2		−0.000	
		[−0.62]	
FOG Q2			0.003**
			[1.99]
FOG Q3			0.006***
			[3.33]
FOG Q4			0.007***
			[3.51]
LOGSIZE	0.049***	0.049***	0.049***
	[65.99]	[66.03]	[66.02]
GROWTH	0.012***	0.011***	0.011***
	[5.54]	[5.52]	[5.50]
LOGSEGMENTS	−0.011***	−0.011***	−0.011***
	[−6.47]	[−6.47]	[−6.48]
PINST	0.000***	0.000***	0.000***
	[10.09]	[10.09]	[10.12]
MFCOUNT	0.003***	0.003***	0.003***
	[8.70]	[8.70]	[8.72]
10-K NEWS	0.049***	0.049***	0.049***
	[4.34]	[4.33]	[4.30]
ADV	0.089***	0.090***	0.089***
	[2.65]	[2.65]	[2.64]
R&D	0.006	0.005	0.005
	[0.92]	[0.89]	[0.83]
STD_RET	0.036***	0.036***	0.036***
	[4.41]	[4.41]	[4.42]
Time Fixed Effects	Yes	Yes	Yes
Industry Fixed Effects	Yes	Yes	Yes
n	30716	30716	30716
Adjusted R^2	0.51	0.51	0.51

**, *** $p < 0.05$, and $p < 0.01$, respectively, two-tailed t-tests.

This table reports coefficient estimates and t-statistics (in brackets) from the regression of the information content of analyst reports on *FOG* and control variables. Analysts information content is defined as the proportion of firm stock returns related to analyst forecast revisions to the total firm stock return during the time period between the 10-K filing and the subsequent fiscal year-end. *FOG QX* is an indicator variable set to 1 if the firm's 10-K Fog Index is in quartile *X* of the sample, and 0 otherwise. All other variables are as defined in Table 2. Industry fixed effects are based on the Fama and French 48-industry classification. t-statistics are robust to heteroscedasticity and clustered at the firm level.

information content of analyst reports by about 0.29 percent to 0.46 percent, on average, across the three models. Consistent with our finding for analyst following, we find that the effect of an

interquartile change in disclosure readability (*FOG*) is similar to many of the other variables in our analysis.[27]

Properties of Analyst Earnings Forecasts

Given our findings regarding the association of disclosure readability with analyst following, analyst report duration, and the information content of analyst reports, a reasonable next step is to examine its association with properties of analyst forecasts. In particular, we focus on the dispersion, accuracy, and the overall and common uncertainty associated with analyst forecasts of annual earnings. As previously explained, we hypothesize that less readable disclosures will be associated with greater forecast dispersion and lower analyst forecast accuracy. We also predict that overall analyst uncertainty will be higher for firms with less readable disclosures, but make no directional prediction regarding analyst common uncertainty.

For each of these hypotheses, we estimate a model of the following form:

$$\begin{aligned}\text{Forecast Property}_{i,t} = {} & \beta_0 + \beta_1 FOG_{i,t} + \beta_2 LOGSIZE_{i,t} + \beta_3 GROWTH_{i,t} + \beta_4 LOGSEGMENTS_{i,t} \\ & + \beta_5 PINST_{i,t} + \beta_6 MFCOUNT_{i,t} + \beta_7 10\text{-}K\ NEWS_{i,t} + \beta_8 ADV_{i,t} \\ & + \beta_9 R\&D_{i,t} + \beta_{10} STD\_RET_{i,t} + \varepsilon_{i,t}, \end{aligned} \quad (6)$$

where "Forecast Property" represents the previously defined variable of interest (*DISPERSION*, *ACCURACY*, $UNCERTAINTY_{OVERALL}$, or $UNCERTAINTY_{COMMON}$). Each model is estimated using ordinary least-squares regression with industry and time fixed effects. Table 7 presents the results. t-statistics, presented in brackets, are based on standards errors that are robust to heteroscedasticity and clustered at the firm level.

As reported in the first two columns of Table 7, we find that the coefficient on *FOG* is positive and significant, indicating that less readable disclosures are associated with more dispersed and less accurate analyst earnings forecasts. Consistent with prior research, these results, in conjunction with our findings about the information content of analysts' reports, suggest that analysts provide information that investors find valuable at the cost of less accurate forecasts (Schipper 1991; Clement and Tse 2003). It is, however, difficult to assess the significance of these benefits because we cannot observe the appropriate benchmark—i.e., the effect on investors of analysts choosing not to follow firms with less readable disclosures.[28]

We further examine these results by testing the effect of 10-K readability on measures of overall and common analyst forecast uncertainty, as derived by Barron et al. (1998). We report our results in columns three and four of Table 7. Consistent with our prediction, we find that overall analyst uncertainty is increasing in *FOG*, suggesting that there is higher overall uncertainty in the analyst information environment for firms with less readable 10-K filings. Further, we find that the proportion of common analyst uncertainty to overall uncertainty is also increasing in *FOG*, suggesting that as readability decreases, publicly available information, such as the 10-K, becomes

[27] As an additional sensitivity test, we follow the suggestion of a reviewer and examine the effect of analysts' conflicts of interest associated with the Global Research Analyst Settlement to support our inferences. Because analysts with conflicts of interest are likely to place less importance on firm disclosures, we predict that our results for analyst following and report information content should be weaker when analysts' conflicts of interest are higher. We note, however, that other important events such as Regulation Fair Disclosure occurred around this time and could limit our inferences from this test. In untabulated results, we find that, consistent with our prediction, the association between analyst following and analyst report information content is stronger in the years following the Global Research Analyst Settlement, suggesting greater analyst reliance and attention to 10-K filings in the post period. This test further supports our overall conclusion regarding the association between analyst behavior and 10-K readability.

[28] Related to this point, Lee (2010) finds that less readable 10-Q filings are associated with lower information efficiency and greater information asymmetry. She also finds that these negative consequences of lower readability are mitigated by analyst coverage, which provides evidence of the benefits of analysts following firms with less readable reports.

TABLE 7

The Association between Annual Report Readability and Properties of Analysts' Information Environment

Variable	*DISPERSION*	*ACCURACY*	$UNCERTAINTY_{OVERALL}$	$UNCERTAINTY_{COMMON}$
Intercept	0.0192***	0.2887	−0.0142	−0.0440
	[4.04]	[1.35]	[−0.37]	[−0.56]
FOG	0.0003***	0.0041**	0.0044***	0.0053**
	[3.96]	[2.10]	[2.58]	[2.23]
LOGSIZE	−0.0027***	−0.0315***	−0.0133***	−0.0012
	[−22.00]	[−9.89]	[−7.19]	[−0.43]
GROWTH	0.0008	0.0077	−0.0005	0.0638***
	[1.08]	[0.45]	[−0.04]	[5.69]
LOGSEGMENTS	0.0008***	0.0134**	0.0025	0.0081
	[3.74]	[1.96]	[0.54]	[1.24]
PINST	−0.0001***	0.0001	−0.0001	0.0010***
	[−8.45]	[0.49]	[−0.77]	[5.48]
MFCOUNT	−0.0001***	0.0010	−0.0001	0.0034***
	[−2.92]	[1.10]	[−0.11]	[2.79]
10-K NEWS	0.0511***	0.7263***	0.5591***	0.3838***
	[7.48]	[5.80]	[4.75]	[4.37]
ADV	0.0266***	0.3901**	0.0933	0.0255
	[4.13]	[2.13]	[1.03]	[0.18]
R&D	0.0081***	−0.0266	−0.0036	−0.1394***
	[5.97]	[−0.87]	[−0.14]	[−4.39]
STD_RET	0.0242***	0.3793***	0.2209***	0.3683***
	[9.27]	[6.03]	[4.36]	[6.72]
Time Fixed Effects	Yes	Yes	Yes	Yes
Industry Fixed Effects	Yes	Yes	Yes	Yes
n	26078	29055	17241	17222
Adjusted R^2	0.20	0.05	0.04	0.06

**, *** $p < 0.05$, and $p < 0.01$, respectively, two-tailed t-tests.

This table reports coefficient estimates and t-statistics (in brackets) from regressions of analyst earnings forecast dispersion, accuracy, and uncertainty on *FOG* and control variables. Analyst earnings forecast dispersion is defined as the standard deviation of the individual analyst earnings forecasts in the most recent consensus earnings forecast following the annual report scaled by price. Analyst earnings forecast accuracy is defined as the squared value of the difference between the I/B/E/S actual reported earnings and the most recent analyst consensus earnings forecast following the annual report filing scaled by price. Analyst overall and common uncertainty are calculated following the equations derived by Barron et al. (1998). All other variables are as defined in Table 2. Industry fixed effects are based on the Fama and French 48-industry classification. t-statistics are robust to heteroscedasticity and clustered at the firm level.

more important to analysts relative to private information. This suggests that analysts have difficulty producing private information for firms with less readable reports. Overall, the evidence from the four regressions in Table 7 suggests that analyst earnings forecasts are affected by 10-K readability, supporting the notion that analysts use the information contained in the 10-K filings.

The results for the control variables are largely consistent with the notion that analysts perform better for firms with better information environments and less complex operations. We find that firms with better information environments, as measured by size and institutional ownership, are associated with greater accuracy, lower dispersion, and lower uncertainty. Institutional ownership is also positively associated with analyst common uncertainty, consistent with these inves-

tors improving the importance of public information. We find that firms with greater uncertainty and complexity, as measured by the volatility of returns, the number of business segments, and the extent of intangible investment, are generally associated with lower forecast accuracy, greater dispersion, and greater overall uncertainty. However, their relation to analyst common uncertainty is less clear. While *R&D* investment is negatively related to common uncertainty (because less public information on these investments is available), return volatility is positively related. These findings suggest that private information is less important relative to common information for firms with high return volatility. While the measures of firm disclosure, *MFCOUNT* and *10-K NEWS*, are both positively associated with analyst common uncertainty (i.e., public disclosures increase the importance of common information), they differ with respect to analyst forecast dispersion. *MFCOUNT* is associated with lower forecast dispersion, consistent with managers providing guidance to reduce analyst disagreement, while *10-K NEWS* is associated with greater dispersion and lower accuracy.

V. SUMMARY AND CONCLUSION

From the passage of the Securities Acts in 1933 and 1934 to the present, regulators, legal scholars, and various other parties have weighed in on the debate about the complexity of financial communication to external users. The SEC has gone so far as to require that the prospectuses of all registered public offerings meet the requirements of the Plain English Rules, and has suggested that similar procedures should be applied to other mandatory filings. Former SEC Chairman Christopher Cox has even suggested the use of readability models such as the Fog Index to measure the complexity of financial communication (Cox 2007). However, other users are concerned about the effects that these types of actions could have on the disclosure regime. These parties argue that, at a minimum, these actions are a waste of time and effort, and could lead to a reduction in disclosure as firms are forced to simplify their public communication (e.g., Kripke 1970, 1973; Firtel 1999). This study sheds insight relevant to this debate by examining the effect of readability on one important financial information intermediary—sell-side financial analysts.

We find evidence consistent with the notion that, because less readable firm communication is more costly to process and interpret, investors demand greater amounts of analyst services for firms with less readable communication. We find that analyst following is greater for firms with higher levels of syntactic complexity as measured by the Fog Index. We also find that analysts who cover firms with less readable communication take longer time on average to issue reports in response to 10-K filings. We interpret this evidence as analysts exerting greater effort to cover these firms. In addition, we find that analyst reports of firms with less readable 10-K reports are more informative to investors, but that the earnings forecasts of such firms have greater analyst dispersion or disagreement, are less accurate, and are associated with greater levels of uncertainty.

Our results suggest that analyst behavior is related to the readability of firms' communication. While prior studies have found that analysts are affected by the complexity of individual financial items, we provide evidence that the *overall* linguistic complexity of firms' communication incrementally influences analyst behavior over and above the effects of the content of the document (e.g., taxes, interest rates). Finally, our results that analysts provide greater amounts of information to investors for firms with less readable communication and that investors consider this information informative are relevant to the SEC's debate about the intended audience of financial information and the SEC's concerns on the accessibility of these reports. While the SEC moves to reduce the complexity of firms' communication, further research is needed to examine the explicit costs and benefits of such actions. As a final caveat, while the collection of the evidence from our various analyses of analysts' behavior conducted in this study is consistent with the prediction that 10-K readability influences analyst behavior, our individual findings still are subject to the limitation that the documented relations reflect associations and may not be fully causal.

REFERENCES

Armstrong, J. S. 1980. Unintelligible management research and academic prestige. *Interfaces* 10 (2): 80–86.

Ball, R. 1992. The earnings price anomaly. *Journal of Accounting and Economics* 15 (2–3): 319–345.

Barron, O. E., D. Byard, O. Kim, and E. J. Riedl. 2002. High-technology intangibles and analysts' forecasts. *Journal of Accounting Research* 40 (2): 289–312.

———, ———, O. Kim, and D. E. Stevens. 1998. Using analysts' forecasts to measure properties of analysts' information environment. *The Accounting Review* 73 (4): 421–433.

Barth, M. E., R. Kasznik, and M. F. McNichols. 2001. Analyst coverage and intangible assets. *Journal of Accounting Research* 39 (1): 1–34.

Bhushan, R. 1989. Firm characteristics and analyst following. *Journal of Accounting and Economics* 11 (2–3): 255–274.

Biddle, G. C., G. Hilary, and R. S. Verdi. 2009. How does financial reporting quality relate to investment efficiency? *Journal of Accounting and Economics.* 48 (2–3): 112–131

Bloomfield, R. J. 2002. The "incomplete revelation hypothesis" and financial reporting. *Accounting Horizons* 16 (3): 233–243.

Botosan, C. A., and M. S. Harris. 2000. Motivations for a change in disclosure frequency and its consequences: An examination of voluntary quarterly segment disclosures. *Journal of Accounting Research* 38 (2): 329–353.

Bradshaw, M. T., G. S. Miller, and G. Serafeim. 2008. Accounting method heterogeneity and analysts' forecasts. Working paper, Harvard University and University of Michigan.

Brennan, M. J., and P. J. Hughes. 1991. Stock prices and the supply of information. *The Journal of Finance* 46 (5): 1665–1691.

———, and A. Subrahmanyam. 1995. Investment analysis and price formation in securities markets. *Journal of Financial Economics* 38 (3): 361–381.

Brown, L. D., and S. Sivakumar. 2003. Comparing the value relevance of two operating income measures. *Review of Accounting Studies* 8 (4): 561–572.

Callen, J. L., M. Khan, and H. Lu. 2009. Accounting quality, stock price delay and future stock returns. Working paper, University of Toronto and Massachusetts Institute of Technology.

Charrow, R. P., and V. R. Charrow. 1979. Making legal language understandable—Psycholinguistic study of jury instructions. *Columbia Law Review* 79 (7): 1306–1374.

Chomsky, N. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Clement, M. B., and S. Y. Tse. 2003. Do investors respond to analysts' forecast revisions as if forecast accuracy is all that matters? *The Accounting Review* 78 (1): 227–249.

Core, J. E. 2001. A review of the empirical disclosure literature: Discussion. *Journal of Accounting and Economics* 31 (1–2): 441–456.

Cotter, J., I. Tuna, and P. D. Wysocki. 2006. Expectations management and beatable targets: How do analysts react to explicit earnings guidance? *Contemporary Accounting Research* 23 (3): 593–628.

Courtis, J. K. 1995. Readability of annual reports: Western versus Asian evidence. *Accounting, Auditing & Accountability Journal* 8 (2): 4–17.

Cox, C. 2007. Closing remarks to the Second Annual Corporate Governance Summit. Delivered at the USC Marshall School of Business, Los Angeles, CA, March 23. Available at http://www.sec.gov/news/speech/2007/spch032307cc.htm.

Duru, A., and D. M. Reeb. 2002. International diversification and analysts' forecast accuracy and bias. *The Accounting Review* 77 (2): 415–433.

Firtel, K. B. 1999. Plain English: A reappraisal of the intended audience of disclosure under the Securities Act of 1933. *Southern California Law Review* 72 (2–3): 851–897.

Frankel, R., S. P. Kothari, and J. Weber. 2006. Determinants of the informativeness of analyst research. *Journal of Accounting and Economics* 41 (1–2): 29–54.

Gazmararian, J. A., D. W. Baker, and M. V. Williams. 1999. Health literacy among Medicare enrollees in a managed care organization. *Journal of the American Medical Association* 281 (6): 545–551.

Gibson, E. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition* 68 (1): 1–76.

Gu, Z. Y., and T. Chen. 2004. Analysts' treatment of nonrecurring items in street earnings. *Journal of Accounting and Economics* 38 (1–3): 129–170.

———, and W. M. Wang. 2005. Intangible assets, information complexity, and analysts' earnings forecasts. *Journal of Business Finance & Accounting* 32 (9–10): 1673–1702.

Gunning, R. 1969. The Fog Index after twenty years. *Journal of Business Communication* 6 (2): 3–13.

Healy, P., 1977. Can you understand the footnotes to financial statements? *Accountants Journal* July: 219–222.

Healy, P. M., A. P. Hutton, and K. G. Palepu. 1999. Stock performance and intermediation changes surrounding sustained increases in disclosure. *Contemporary Accounting Research* 16 (3): 485–520.

Hirst, D. E., and P. E. Hopkins. 1998. Comprehensive income reporting and analysts'' valuation judgments. *Journal of Accounting Research* 36: 47–75.

———, ———, and J. M. Wahlen. 2004. Fair values, income measurement, and bank analysts' risk and valuation judgments. *The Accounting Review* 79 (2): 453–472.

Hong, H., and J. D. Kubik. 2003. Analyzing the analysts: Career concerns and biased earnings forecasts. *The Journal of Finance* 58 (1): 313–351.

Hopkins, P. E., R. W. Houston, and M. F. Peters. 2000. Purchase, pooling, and equity analysts' valuation judgments. *The Accounting Review* 75 (3): 257–281.

Indjejikian, R. J. 1991. The impact of costly information interpretation on firm disclosure decisions. *Journal of Accounting Research* 29 (2): 277–301.

Johnston, R., S. Markov, and S. Ramnath. 2009. Sell-side debt analysts. *Journal of Accounting and Economics* 47: 91–107.

Jones, M. J., and P. A. Shoemaker. 1994. Accounting narratives: A review of empirical studies of content and readability. *Journal of Accounting Literature* 13: 142–184.

Kimble, J. 1994. Answering the critics of plain language. *The Scribes Journal of Legal Writing* 5: 51–85.

Koo, M. M., I. Krass, and P. Aslani. 2003. Factors influencing consumer use of written drug information. *The Annals of Pharmacotherapy* 37 (2): 259–267.

Kripke, H. 1970. The SEC, the accountants, some myths and some realities. *New York University Law Review* 45 (6): 1151–1205.

———. 1973. The myth of the informed layman. *Business Law* 28: 631–638.

Lang, M. H., and R. J. Lundholm. 1996. Corporate disclosure policy and analyst behavior. *The Accounting Review* 71 (4): 467–492.

———, K. V. Lins, and D. P. Miller. 2004. Concentrated control, analyst following, and valuation: Do analysts matter most when investors are protected least? *Journal of Accounting Research* 42 (3): 589–623.

Lee, S. Y. D., J. A. Gazmararian, and A. M. Arozullah. 2006. Health literacy and social support among elderly Medicare enrollees in a managed care plan. *Journal of Applied Gerontology* 25 (4): 324–337.

Lee, Y. 2010. The effects of quarterly reported readability on information efficiency of stock prices. Working paper, National Taiwan University.

Leuz, C., and C. M. Schrand. 2009. Disclosure and the cost of capital: Evidence from firms' response to the Enron shock. Working paper, University of Chicago and University of Pennsylvania.

Li, F. 2008. Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics* 45 (2–3): 221–247.

Li, Y., P. R. Rau, and J. Xu. 2009. All-star analyst coverage and corporate transparency. Working paper, Purdue University.

Loughran, T., and B. McDonald. 2010. Measuring readability in financial text. Working paper, University of Notre Dame.

Mikhail, M. B., B. R. Walther, and R. H. Willis. 1999. Does forecast accuracy matter to security analysts? *The Accounting Review* 74 (2): 185–200.

Miller, P. B. 2010. The effects of reporting complexity on small and large investor trading. *The Accounting Review* 85 (6): 2107–2143.

Nagar, V., D. Nanda, and P. Wysocki. 2003. Discretionary disclosure and stock-based incentives. *Journal of Accounting and Economics* 34 (1–3): 283–309.

O'Brien, P. C., M. F. McNichols, and H. W. Lin. 2005. Analyst impartiality and investment banking relationships. *Journal of Accounting Research* 43 (4): 623–650.

———, and R. Bhushan. 1990. Analyst following and institutional ownership. *Journal of Accounting Research* 28: 55–76.

Ott, B. B., and T. L. Hardie. 1997. Readability of advance directive documents. *Image—the Journal of Nursing Scholarship* 29 (1): 53–57.

Phillips, T. J., C. M. Daily, and M. S. Luehlfing. 2007. A note on the readability of professional materials for management accountants. *Advances in Management Accounting* 16: 311–318.

Plumlee, M. A. 2003. The effect of information complexity on analysts' use of that information. *The Accounting Review* 78 (1): 275–296.

Previts, G. J., R. J. Bricker, T. R. Robinson, and S. J. Young. 1994. A content analysis of sell-side financial analysts company reports. *Accounting Horizons* 8 (2): 55–70.

Rogers, R. K., and J. Grant. 1997. Content analysis of information cited in reports of sell-side financial analysts. *Journal of Financial Statement Analysis* 3 (1): 17–30.

Schipper, K. 1991. Analysts' forecasts. *Accounting Horizons* 5 (4): 105–121.

Schroeder, M. 2002. SEC proposes rules to improve disclosure by public companies. *Wall Street Journal* (May 1): C5.

Schroeder, N. and C. Gibson. 1990. Readability of management's discussion and analysis. *Accounting Horizons* 4 (4): 78–87.

Securities and Exchange Commission (SEC). 1998. *A Plain English Handbook: How to Create Clear SEC Disclosure Documents*. Washington, D.C.: U.S. Securities and Exchange Commission.

Shuptrine, F. K., and E. M. Moore. 1980. Even after the Magnuson-Moss Act of 1975, warranties are not easy to understand. *The Journal of Consumer Affairs* 14 (2): 394–404.

Smith, J. E., and N. P. Smith. 1971. Readability: A measure of the performance of the communication function of financial reporting. *The Accounting Review* 46 (3): 552–561.

Stone, C. A., E. R. Silliman, B. J. Ehren, and K. Apel. 2005. *Handbook of Language and Literacy: Development and Disorders*. New York, NY: Guilford Press.

Thompson, C. K., and Y. Faroqi-Shah. 2002. Models of sentence production. In *Handbook of Adult Language Disorders: Integrating Cognitive Neuropsychology, Neurology, and Rehabilitation*, edited by A. E. Hillis, 311–330. Philadelphia, PA: Psychology Press, Taylor and Francis.

You, H., and X. Zhang. 2009. Financial reporting complexity and investor underreaction to 10-K information. *Review of Accounting Studies* 14 (4): 559–586.

THE ACCOUNTING REVIEW
Vol. 86, No. 3
2011
pp. 1117–1122

American Accounting Association
DOI: 10.2308/accr.00000035

BOOK REVIEWS

Stephen A. Zeff, Editor

Editor's note: Two copies of books for review should be sent to the Book Review Editor: Stephen A. Zeff, Rice University, Jesse H. Jones Graduate School of Business, 6100 Main St., Houston, TX 77005. The policy of *The Accounting Review* is to publish only those reviews solicited by the Book Review Editor. Unsolicited reviews will not be accepted.

JOAN LUFT and MICHAEL D. SHIELDS, *Psychology Models of Management Accounting*, Foundations and Trends® in Accounting (Hanover, MA: now Publications, Inc., 2009, ISBN 978-1-60198-346-6, Vol. 4, No. 3–4, pp. 151).

This monograph is the newest in the series entitled "Foundations and Trends® in Accounting." The objective of the series is to provide frameworks for understanding and organizing the vast array of existing literature in various fields of accounting research. This issue is co-authored by Joan Luft and Mike Shields, two thought leaders in the field of psychology-based research in management accounting. The main objective of the monograph is to devise a framework that links what we know about the subjective decisions made by management accountants with the theories in psychology that have proven useful to our understanding of how these decisions are made. These authors have previously developed a comprehensive study of the causes and consequences of management accounting based on an analysis of 275 papers across six leading journals (see Luft and Shields 2003). The current monograph provides a framework for understanding a much more specific and more focused piece of the management accounting literature, that is, the psychological processes underlying the subjective decisions required of management accountants in practice.

Given the vast array of possible pieces of the literature to review, the reader might ask, "Why is this book needed now?" Interest in the field of behavioral decision research in accounting has spiked due at least in part to the awarding of the Nobel Prize in Economic Sciences in 2002 to Daniel Kahneman and Amos Tversky for their work on the psychology of judgment and decision-making and its application to the field of behavioral economics. Although much psychology-based research in management accounting took place before 2002, the recognition of behavioral economics as an important economic discipline seems to have moved behavioral decision theory in management accounting to the forefront. The literature reviewed in this monograph reflects the best of what has been published in this area since then.

This monograph should prove useful to management accounting scholars and doctoral students interested in behavioral decision theory and its applications in management accounting. Doctoral students would benefit from the authors' framework for thinking about subjective decision making in management accounting. In addition, the organized framework allows scholars to more easily identify gaps in the literature that they may wish to fill.

Do not be confused, as this not a one-to-one mapping of decision-making categories to psychological theories; in fact, some psychological theories are discussed as underpinnings of several of the decision categories identified by the authors. Such confusion may be driven in part by the title—*Psychology Models of Management Accounting*. The authors do discuss models, but I would argue they are management accounting models of subjective decision making. I might re-title the monograph "How Psychological Theory Informs our Understanding of Subjective Decision Making in Management Accounting."

The monograph is structured in seven chapters plus an appendix. The authors have a clear and concise writing style reflective of many years of academic writing experience. Chapter 1 provides an introduction to the monograph's purpose and objectives as well as an overview of the authors' organizing framework for subjective decision making in management accounting. The organizing framework consists of four categories of subjective

decision making in management accounting and two emerging themes. The authors argue that there is robust evidence consistent with psychological theory underlying the four categories, while there is some interesting, but perhaps less convincing (or at least less well developed), evidence to support the two emerging themes.

Chapter 2 is meant to describe the methodology used by the authors to develop their categorization scheme. The most important information which the reader needs to understand, namely the methodology, is unfortunately relegated to the appendix and the last paragraph of Chapter 7. The authors instead devote much of Chapter 2 to an argument in favor of cognitive task analysis (i.e., a description and analysis of the knowledge and cognitive activities required for high performance on specific tasks) in management accounting. The authors argue that we have not made a point as a discipline to perform detailed cognitive task analyses of the type reported in the audit literature. Consequently, any framework for understanding how psychology affects management accounting decisions must be organized in an alternate form. Although this issue seems somewhat beside the point in the context of the chapter material, the authors' argument concerning the importance of cognitive task analysis (or lack thereof) in management accounting is well taken. The field would benefit from a knowledge of the detailed cognitive processes involved in subjective decision making in management accounting.

At this point, the reader should be aware of the importance of the appendix in order to obtain a good understanding of the authors' methodology. The appendix walks the reader through two examples, one decision-facilitating and one decision-influencing, of the use of management accounting information in subjective decision making. The decision-facilitating example is based on the "grand model" of product capacity planning and pricing developed by Balakrishnan and Sivaramakrishnan (2002). The decision-influencing example references the work of Feltham and Xie (1994), who model the weights placed on performance measures in a multi-measure performance evaluation scheme. In both cases, the authors point out the many individual (subjective) decisions which managers must make in the course of capacity planning and performance evaluation. The implication is that the authors grouped the various individual decisions made in the process of capacity planning and performance evaluation into the categories listed in Chapter 2.

While the examples provided in the appendix are somewhat informative, the link between the examples and the categories derived from them is rather indirect. The authors are not particularly forthcoming about how the appendix maps onto the categories; consequently, the reader must infer the mapping. While this would not be too difficult for experienced management accounting scholars, doctoral students or non-management accounting researchers would have a more difficult time understanding how the lists provided in the appendix distill down to the four categories identified by the authors in developing their framework, potentially limiting its usefulness.

Chapters 3–6 are devoted to discussions of each of the four subjective decision-making categories. For each category, the authors describe the important elements of the decision process and identify relatively recent studies that examine the type of decision under consideration. The authors do a nice job of summarizing each paper and its potential contribution as well as how later papers build on earlier papers to fill the gaps in our understanding. Chapter 3 examines subjective valuation of management accounting-related monetary payoffs, linking these mainly to cognitive psychology, including framing and prospect theory. Chapter 4 examines subjective valuation of non-monetary payoffs, including fairness, honesty, trust, and reciprocity, mainly supported by theories of social and relational exchange drawn from social psychology. Chapter 5 examines the relatively simple subjective decision models used to perform many management accounting-related tasks (e.g., subjective linear programming, mental models, non-compensatory decision models, effects of opportunity, and sunk costs), while Chapter 6 examines models of parameter and variable acquisition and subjective estimation (e.g., subjective estimation of model parameters using accounting information, effect of accounting knowledge, and experience on this subjective estimation). Each of Chapters 3–6 also includes an exhibit indicating the link between causes (i.e., management accounting independent variables) and effects (i.e., management accounting dependent variables) through particular psychological theories. In addition, the exhibits indicate which of the management accounting studies reviewed in the chapter address the links identified in the exhibit, thus providing nice visual summaries of the main points raised in each of these chapters.

After reading these chapters, I wondered how the particular illustrative articles were selected by the authors for inclusion in the monograph. Most of the illustrative articles were published since 2000, and all were published in top-tier accounting journals. The list of articles included in the monograph is by no means exhaustive, nor do I believe the authors meant it to be. The authors justify their focus on recent work because it reflects the "growth and dynamism of psychology-based research" on management accounting (p. 124). This seems reasonable. In addition, I suspect the authors selected the papers based on their extensive collective experience teaching and researching subjective decision making in management accounting.

Finally, Chapter 7 discusses the two identified emerging themes to motivate future work. The emerging themes are (1) that two or three distinctly different subjective models account for most management accounting decisions, and (2) the important differences between intuitive or "gut-feel" decision making (or, as psychologists have labeled them, "System 1 cognitions") and deliberative or consciously controlled decision making (i.e., "System 2 cognitions") and implications of these for management accounting. According to the authors, these areas are ripe for future research. In addition, the authors identify several important, but relatively unexplored, research questions that could benefit from the application of insights from psychology, including:

- Why not study larger, more complex decision models that better reflect real-world decision making?
- What drives selective information acquisition, given that modern information systems can provide so many different pieces of information?
- Why don't we see more non-laboratory studies similar to Ittner et al. (2003) to examine management accounting subjective decision making in real-world settings?

In conclusion, I believe this book provides a good primer for scholars new to management accounting behavioral research on subjective decision making. For scholars already working in the field, the authors provide an organizing framework for thinking about the vast array of work in this area. This allows us to figure out where our research fits within the current literature and to pinpoint areas where future work is needed to move the literature ahead.

For many years, we have talked about integrating economics and psychology-based research in management accounting. Because of the increased level of interest in behavioral economics and behavioral finance, many economics-based researchers are becoming more interested in identifying behavioral theories that would allow them to gain additional insight into their research questions of interest. My hope is that this increased interest leads to fruitful partnerships between economics- and psychology-based researchers in management accounting. In my opinion, this monograph provides a much-needed tool to inform economics-based researchers of the "state of the art" in subjective decision-making research in management accounting. Consequently, this monograph should be helpful in further encouraging such partnerships.

REFERENCES

Balakrishnan, R., and K. Sivaramakrishnan. 2002. A critical overview of the use of full-cost data for planning and pricing. *Journal of Management Accounting Research* 14: 3–31.

Feltham, G., and J. Xie. 1994. Performance measure congruity and diversity in multi-task principal/agent relations. *The Accounting Review* 69 (3): 429–453.

Ittner, C. D., D. F. Larcker, and M. W. Meyer. 2003. Subjectivity and the weighting of performance measures: Evidence from a balanced scorecard. *The Accounting Review* 78 (3): 725–758.

Luft, J., and M. D. Shields. 2003. Mapping management accounting: Graphics and guidelines for theory-consistent empirical research. *Accounting, Organizations and Society* 28: 169–249.

THERESA LIBBY
Professor of Accounting
University of Waterloo

CHRIS POULLAOS and SUKI SIAN (editors), *Accountancy and Empire: The British Legacy of Professional Organization* (New York, NY: Routledge, 2010, ISBN 978-0-415-45771-2, pp. xv, 263).

This book discusses the historical development of the accounting profession in different parts of the British Empire. Its objective is to address a gap in the literature "by tracing the rise of professional accountancy in selected colonial outposts and examining the role of the institutions regulating it, within the context of the empire experience" (p. 2). It includes contributions of expert academics, some of them having published extensively on the development of accounting professions in different contexts, others focusing their research on the region that they discuss in their contribution to this work. Some of the chapters build upon previously published work, whereas others try to explore new aspects of the history of accountancy. Some of the countries that are discussed in this book are also found in Poullaos (2009), which can serve as a more general introduction to the history of accountancy as a profession.

Although the different chapters clearly focus on the same type of issues, their development is rather variable. This is partly a consequence of the different historical methodologies on which they are based. In Chapter 2 (on South Africa, Australia, and Canada), Chris Poullaos has relied on an extensive set of primary sources from, e.g., the ICAEW archives, the National Archives, and the Scottish Records Office. It is the only contribution that relies heavily on unpublished archival sources. Suki Sian (Chapter 10, on Kenya) and Shraddha Verma (Chapter 9, on India) use a mixture of primary and secondary sources. Chibuike U. Uche (Chapter 4, on Nigeria) refers also to

email correspondence with representatives of professional bodies. The contribution of Devi S. Susela (Chapter 5, on Malaysia) is based on 65 interviews, whereas P. W. Senarath Yapa (Chapter 6, on Sri Lanka) uses information from nine unstructured interviews. The other contributions rely mainly on secondary sources.

The book covers a range of countries, including the self-governing (white) settler colonies, South Africa, Australia, and Canada, and non-settler colonies such as Nigeria, Malaysia, Sri Lanka, Jamaica, Trinidad and Tobago, India, and Kenya. Consequently, it covers most regions of the British Empire, the only notable exception being what could be considered the Arab part of the Empire. Although the current selection of former colonies provides contrasting examples of the uptake of the British models, it would be interesting to study whether the situation in protectorates such as British Somaliland, Yemen, and Oman was any different from that in the other parts of the Empire.

The subtitle of the book, "*The British Legacy of Professional Organization*," is highly relevant, as it clarifies that the focus of the contributions is on the organization of accountancy in occupational groupings, not on the development of accounting and auditing practices. Only in some chapters is attention given to the development of individual audit firms, most notably in the chapter on Trinidad and Tobago. All chapters provide some background on the countries involved. For most of them, the information provided is sufficient to understand the issues at stake, but in some chapters more circumstantial information would make the reading easier. Most chapters center strongly on the accounting profession, and one wonders whether similar discussions and developments could be found in other professions.

In most of the chapters, the key issue is the development of a local accounting profession and the relation between British accountancy bodies and overseas associations. An overall impression after reading the book is that this was to a large extent a one-way relation, with, in a first stage, British accountants being "exported" to the different parts of the Empire and, in a later phase, locally trained accountants who organized themselves, very often following British models and with direct or indirect interventions by the U.K. bodies. Even the competition that existed between the U.K. bodies in their home market had a clear impact on developments overseas. This is also apparent in the concluding section, where Poullaos and Sian address how the organization of accountants in different parts of the empire has been affected by the empire experience. This section also focuses on the relation between British capital and the development of colonial professional organization. It explains how South Africa, Australia, and Canada were "decolonized" earlier than the other colonies in terms of formal control over accountancy. In many of the non-settler colonies, local legislation confirmed the continued impact of British reporting, auditing, and professional requirements. It led to professional exclusion processes both before and after political independence. In these countries, local accountants typically could organize themselves only after decolonization. The closing section also discusses how the empire experience had an impact on the organization of accountancy in Britain. However, as the different chapters show, this impact seems limited: contrary to the many British expatriate accountants who went to, for example, Kenya, the opposite flow of trainees was very small. An interesting extension of the current research might focus on analyzing quantitative data on, e.g., the number of non-British citizens who were members of the Institute of Chartered Accountants in England and Wales and other U.K. bodies (see also Briston and Kedslie 1997). Some chapters give the impression that they were very rare (e.g., the chapter on Nigeria), but others indicate that there were "many locals" who went to the U.K. for professional qualification (e.g., the chapter on Malaysia).

Apart from the introductory and concluding chapters, each chapter focuses on a group of related colonies or dominions or on an individual country. Poullaos discusses the evolution of the imperial accountancy arena during the 1920s, focusing on South Africa, Australia, and Canada. He extends the earlier work of Chua and Poullaos (2002) on the interactions between British accountants and their associations on the one hand, and dominion accountants on the other hand. Issues at stake include the use of the "chartered accountant" designation, access by British accountants to dominion markets and of dominion accountants to British designations, and participation in British examinations without having to travel to Britain. The chapter provides a quite detailed account of initiatives, events, and negotiations and demonstrates how, by 1930, South African, Canadian, and Australian accountants in different ways had strengthened their position *vis-à-vis* the British profession.

Canada is also the subject of the second chapter, but Alan J. Richardson draws attention to the fact that the British Empire was not the only imperial power that had an impact on the development of accountancy in Canada. He argues that the Canadian profession was shaped by French, British, and American institutional legacies and that the impact of each of these changed over time. These competing influences still exist today, and their inconsistencies allowed Canada to go its own way. Changing professional designations (chartered accountants, certified public accountants …) exemplify different paths in the development of the profession. Although, overall, the American influence increased, this influence seems to have been more on the standard-setting processes than on the development of professional structures. The chapter ends with a number of examples in which Richardson shows Canada's position as a "middle power" with a distinct point of view on many issues.

The chapter on Jamaica makes it clear that the power struggle between the U.K. institutes was exported to the colonies. This is most obvious in the opposition between the Association of Certified Corporate Accountants (as

the ACCA was then known) and the other bodies. Different professional bodies coexisted, some linked with U.K. bodies, others being led by indigenous accountants. Again, professional designations were important, with conflicts between "chartered" and "certified" accountants that were often related to attempts to become more independent of the British accounting bodies. Specific issues that are discussed in the chapter are the attempt to create an accounting degree at the University of the West Indies that would serve as an entry qualification for the profession, and the entry of the American Institute of Certified Public Accountants. The ACCA played a similar role in Trinidad and Tobago (T&T). Marcia Annisette's contribution focuses on the evolution of the practice link between Britain and T&T. She starts from the dominance of the large multinational audit firms in the 1990s and traces their history in T&T. There were two different phases in the internationalization of accounting practice, the first one linked with colonialism and British accountancy, the second one in the 1960s linked with U.S.-based multinational practices. This increasing impact of international networks should be seen in a context of growing international standardization of accountancy practices.

There are two contributions on African countries, in which Uche discusses Nigeria and Sian treats Kenya. Uche observes that, around the mid 20th century, Nigeria's export trade was dominated by European firms and that expatriate (U.K.) accountants were the only ones present in the country. This colonial influence still continues and impedes training second-tier accountants who could provide services for smaller local business. Only with the coming of independence did Nigerians become qualified as accountants. It is interesting to note that, again, the ACCA played a major role in this, probably due to the structure of its training programs. Uche reports in detail the continuing impact of the ACCA and its interactions with the formation of local professional bodies. In a similar way, Kenya could not break radically with its colonial past following independence in 1963. Its situation was quite comparable to that of Nigeria, a difference being a stronger economic influence of Asians in the former. Sian suggests that the creation of the Association of Accountants in East Africa, acting as an interest group for British accountants, might have been an attempt to close off the market for non-whites. After independence, movements toward an indigenous profession remained influenced by British concepts of professionalism and British imperial history, although eventually an autonomous profession emerged.

Asian countries included are Malaysia, Sri Lanka, and India. Also in these countries, the continuing impact of British accountants and resistance against this influence had a profound impact on the development of accountancy. The development of the profession in Malaysia is characterized by the complex relation between the Malaysian Association of Certified Public Accountants, the Malaysian Institute of Accountants (a government body), and British professional bodies. This relation seems quite interesting, and it shows some specific characteristics. One of these is that Malaysia was not allowed to use the CA designation, a point that is not fully explained in the chapter. The situation in Sri Lanka seems at first rather comparable, but then again different. Competition there was mainly between the Institute of Chartered Accountants of Sri Lanka (ICASL, having members with British qualifications) and the Registered Accountants (local graduates, considered to be the lower tier of the profession). The ICASL managed to reserve the public practice market to a limited number of elite accountants, a closure that it could maintain until 1977. The Indian professional model was also based on the British model. However, interventions by the Indian government were significant, and they resulted in changes in the regulation of the profession. The chapter provides an interesting summary of this interaction between state and profession. Also here, the use of the "chartered" designation was an issue and seems linked with the social stratification within the profession.

The issue of the designation of accounting professionals can perhaps be seen as a red thread throughout the book. Many debates on the relation between British accountants, British accountancy bodies, and local professionals seem to translate into claims on the "Chartered Accountant" designation that apparently served as a quality label (see also Parker 2005).

In conclusion, this book provides an interesting overview of how professional structures and concepts that were introduced under the rule of the British Empire continued to have an impact after the independence of the countries involved. For scholars who want to understand the current status and organization of the accounting profession in these countries, this book is essential reading. Generally speaking, the way these professions are organized is shaped both by a continuing colonial legacy and by the way local governments and professionals have interacted with these colonial remains in order to create an autonomous profession of indigenous accountants.

REFERENCES

Briston, R. J., and M. J. M. Kedslie. 1997. The internationalization of British professional accounting: The role of the examination exporting bodies. *Accounting Business & Financial History* 7 (2): 175–194.

Chua, W. F., and C. Poullaos. 2002. The Empire strikes back? An exploration of centre-periphery interaction between the ICAEW and accounting associations in the self-governing colonies of Australia, Canada and South Africa, 1880–1907. *Accounting, Organizations and Society* 27 (4/5): 409–445.

Parker, R. H. 2005. Naming and branding: Accountants and accountancy bodies in the British Empire and Commonwealth, 1853–2003. *Accounting History* 10 (1): 7–46.
Poullaos, C. 2009. Professionalisation. In *The Routledge Companion to Accounting History*, edited by Edwards, J. R., and S. P. Walker, 247–273. London, U.K.: Routledge.

IGNACE DE BEELDE
Professor of Accounting
Ghent University and Grenoble Ecole de Management

Capsule Commentary

By the Book Review Editor

DONALD G. TROW and STEPHEN A. ZEFF, *Accounting Education and the Profession in New Zealand: Profiles of the Pioneering Academics and the Early University Accounting Departments 1900–1970* (Wellington, N.Z.: New Zealand Institute of Chartered Accountants, 2010, ISBN 978-1-877529-06-1, pp. iii, 98).

This small monograph contains summary analyses of the early evolution of the accounting departments at Victoria University of Wellington, the University of Otago, the University of Canterbury, and the University of Auckland, based largely on 17 interviews which the authors conducted in 2001. In addition, they provide profiles of seven significant accounting academics and 14 biographical sketches of all of the full professors plus a few other important academics during the era covered by the monograph. There is then a 24-page enumeration of books and articles published by New Zealand accounting academics between 1900 and 1970. Four pages of photographs, on plate stock, of the leading academic figures in the four departments are included as well. Roger W. Hopkins contributed the Foreword.

This is the first published historical record of early New Zealand accounting academe bridging the departments in the original four universities.

EDITORIAL POLICY AND STYLE INFORMATION

EDITORIAL POLICY

According to the policies set by the Publications Committee (which were endorsed by the Executive Committee and were published in the *Accounting Education News*, June 1987), *The Accounting Review* "should be viewed as the premier journal for publishing articles reporting the results of accounting research and explaining and illustrating related research methodology. The scope of acceptable articles should embrace any research methodology and any accounting-related subject, as long as the articles meet the standards established for publication in the journal ... No special sections should be necessary. The primary, but not exclusive, audience should be—as it is now—academicians, graduate students, and others interested in accounting research."

The primary criterion for publication in *The Accounting Review* is the significance of the contribution an article makes to the literature. Topical areas of interest to the journal include accounting information systems, auditing and assurance services, financial accounting, management accounting, taxation, and all other areas of accounting, broadly defined. The journal is also open to all rigorous research methods.

The efficiency and effectiveness of the editorial review process is critically dependent upon the actions of both authors submitting papers and the reviewers. Authors accept the responsibility of preparing research papers at a level suitable for evaluation by independent reviewers. Such preparation, therefore, should include subjecting the manuscript to critique by colleagues and others and revising it accordingly prior to submission. The review process is not to be used as a means of obtaining feedback at early stages of developing the research.

Reviewers and editors are responsible for providing constructive and prompt evaluations of submitted research papers based on the significance of their contribution and on the rigor of analysis and presentation.

MANUSCRIPT PREPARATION AND STYLE

The Accounting Review's manuscript preparation guidelines follow *The Chicago Manual of Style* (15th ed.; University of Chicago Press). Another helpful guide to usage and style is *The Elements of Style*, by William Strunk, Jr., and E. B. White (Macmillan). Spelling follows *Webster's Collegiate Dictionary*.

FORMAT

1. All manuscripts should be formatted in 12-point font on 8 1/2 x 11″ paper and should be double-spaced, except for indented quotations.
2. Manuscripts should be as concise as the subject and research method permit, generally not to exceed 7,000 words.
3. Margins should be at least one inch from top, bottom, and sides.
4. To promote anonymous review, authors should not identify themselves directly or indirectly in their papers or in experimental test instruments included with the submission. Single authors should not use the editorial "we".
5. A cover page should show the title of the paper, all authors' names, titles and affiliations, email addresses, and any acknowledgments.
6. The American Accounting Association encourages use of gender-neutral language in its publications.

Pagination: All pages, including tables, appendices and references, should be serially numbered. Major sections should be numbered in Roman numerals. Subsections should not be numbered.

Numbers: Spell out numbers from one to ten, except when used in tables and lists, and when used with mathematical, statistical, scientific, or technical units and quantities, such as distances, weights and measures. For example: *three days; 3 kilometers; 30 years*. All other numbers are expressed numerically.

Percentages and Decimal Fractions: In nontechnical copy use the word percent in the text.

Hyphens: Use a hyphen to join unit modifiers or to clarify usage. For example: *a cross-sectional equation; re-form*. See *Webster's* for correct usage.

Keywords: The abstract must be followed by at least three keywords to assist in indexing the paper and identifying qualified reviewers.

ABSTRACT/INTRODUCTION

An Abstract of about 100 words (150 maximum) should be presented on a separate page immediately preceeding the text. The Abstract should concisely inform the reader of the manuscript's topic, its methods, and its findings. The Keywords statement should appear immediately below the Abstract. The text of the paper should start with a section labeled "I. Introduction," which provides more details about the paper's purpose, motivation, methodology, and findings. Both the Abstract and the Introduction should be relatively nontechnical, yet clear enough for an informed reader to understand the manuscript's contribution. The manuscript's title, but neither the author's name nor other identification designations, should appear on the Abstract page.

TABLES AND FIGURES

The author should note the following general requirements:

1. Each table and figure (graphic) should appear on a separate page and should be placed at the end of the text. Each should bear an Arabic number and a complete title indicating the exact contents of the table or figure. Tables and figures should define each variable. The titles and definitions should be sufficiently detailed to enable the reader to interpret the tables and figures without reference to the text.
2. A reference to each graphic should be made in the text.
3. The author should indicate where each graphic should be inserted in the text.
4. Graphics should be reasonably interpreted without reference to the text.
5. Source lines and notes should be included as necessary.

Equations: Equations should be numbered in parentheses flush with the right-hand margin.

DOCUMENTATION

Citations: Within-text citations are made using an author-year format. Cited works must correspond to the list of works listed in the "References" section. Authors should make an effort to include the relevant page numbers in the within-text citations.

1. In the text, works are cited as follows: author's last name and year, without comma, in parentheses. For example: one author, (Berry 2003); two authors, (Fehr and Schmidt 2003); three or more authors, (Dechow et al. 1998); more than one work cited, (Cole and Yakushiji 1984; Dechow et al. 1995; Levitt 1998); with two works by the same author(s), (Nelson 2003, 2005).
2. Unless confusion would result, do not use "p." or "pp." before page numbers. For example, (Dechow and Dichev 2002, 41–42).
3. When the reference list contains two or more works by the same author (as the only author or first of two or more authors) published in the same year, the suffix a, b, etc., is appended to the date in the within-text citations and in the "References" section. For example, (Johansson 2004a, 2004b, 2004c; Baiman and Rajan 2002a, 2002b; Dhaliwal et al. 2005a, 2005b).
4. When the author's name is mentioned in the text, it need not be repeated in the citation. For example: "Cohen et al. (2005) provide ..."
5. Citations to institutional works should use acronyms or short titles where practicable. For example: (NCFFR, The Treadway Commission 1987).
6. If the paper refers to statutes, legal treatises, or court cases, citations acceptable in law reviews should be used.

Reference List: Every manuscript must include a "References" section that contains only those works cited within the text. Each entry should contain all information necessary for unambiguous identification of the published work. Use the following formats (which follow *The Chicago Manual of Style*):

1. Arrange citations in *alphabetical order* according to the surname of the first author or the name of the institution or body responsible for the published work.
2. Use authors' initials instead of proper names.
3. For two or more authors, separate authors with a comma, including a comma before "and" (Dechow, P. M., R. Sloan, and A. Sweeney).
4. Date of publication follows the author's (authors') name(s).
5. Titles of journals or newspapers should not be abbreviated.

Sample entries are as follows:

Baiman, S., and M. Rajan. 2002a. The role of information and opportunism in the choice of buyer-supplier relationships. *Journal of Accounting Research* 40 (2): 247–278.

———, and ———. 2002b. Incentive issues in inter-firm relationships. *Accounting, Organizations and Society* 27 (3): 213–238.

Berry, R. 2003. Testimony before the Senate Committee on Homeland Security and Governmental Affairs Permanent Subcommittee on Investigations. November 18. Available at: http://hsgac.senate.gov/files/111803berry.pdf.

Cohen, D., A. Dey, and T. Lys. 2005. The Sarbanes Oxley Act of 2002: Implications for compensation structure and risk-taking incentives of CEOs. Working paper, New York University, University of Chicago, and Northwestern University.

Cole, R., and T. Yakushiji, eds. 1984. *The American and Japanese Auto Industries in Transition.* Ann Arbor, MI: University of Michigan.

Dechow, P. M., R. Sloan, and A. Sweeney. 1995. Detecting earnings management. *The Accounting Review* 70 (2): 193–225.

———, S. P. Kothari, and R. L. Watts. 1998. The relation between earnings and cash flows. *Journal of Accounting and Economics* 25: 133–168.

———, and I. Dichev. 2002. The quality of accruals and earnings: The role of accrual estimation errors, *The Accounting Review* 77 (Supplement): 35–59.

Dhaliwal, D., M. Erickson, and O. Li. 2005a. Shareholder income taxes and the relation between earnings and returns. *Contemporary Accounting Research* 22: 587–616.

———, L. Krull, O. Li, and W. Moser. 2005b. Dividend taxes and implied cost of equity capital. *Journal of Accounting Research* 43: 675–708.

Easton, P. 2003. Discussion of: The predictive value of expenses excluded from pro forma earnings. *Review of Accounting Studies* 8: 175–183.

Engel, E., R. Hayes, and X. Wang. 2007. The Sarbanes-Oxley Act and firms' going-private decisions. *Journal of Accounting and Economics* (forthcoming).

Financial Accounting Standards Board (FASB). 2006. *Accounting for Uncertainty in Income Taxes, and Interpretation of FASB Statement No. 109.* FASB Interpretation No. 48. Financial Accounting Series. Norwalk, CT: FASB.

Fehr, E., and K. Schmidt. 2003. A theory of fairness, competition, and cooperation. In *Advances in Behavioral Economics*, edited by C. Camerer, G. Loewenstein, and M. Rabin, 271–296. New York, NY: Princeton University Press.

Janis, I. L. 1982. *Groupthink: Psychological Studies of Policy Decisions and Fiascoes.* Boston, MA: Houghton Mifflin.

Levitt, A. 1998. The numbers game. Speech delivered at New York University, Center for Law and Business, September 28.

Maggi, G. 1999. The value of commitment with imperfect observability and private information. *RAND Journal of Economics* (Winter) 30: 555–574.

National Commission on Fraudulent Reporting (the Treadway Commission). 1987. *Report of the National Commission on Fraudulent Financial Reporting.* Washington, D.C.: NCFFR.

Nelson, M. W. 2003. Behavioral evidence on the effects of principles- and rules-based standards. *Accounting Horizons* 17 (1): 91–104.

———. 2005. A review of experimental and archival conflicts-of-interest research in auditing. In *Conflicts of Interest: Challenges and Solutions in Business, Law, Medicine, and Public Policy*, edited by D. A. Moore, D. M. Cain, G. Loewenstein, and M. H. Bazerman. Cambridge, U.K.: Cambridge University Press.

Rigdon, E. E., R. E. Schumacker, and W. Wothke. 1998. A comparative review of interaction and nonlinear modeling. In *Interaction and Nonlinear Effects in Structural Equation Modeling*, edited by R. E. Schumacker, and G. A. Marcoulides, 1–16. Mahwah, NJ: Erlbaum Associates.

Schultz, E., and T. Francis. 2002. Companies profit on workers' deaths through "dead peasants" insurance. *Wall Street Journal* (April 19): 1.

Securities and Exchange Commission (SEC). 2002. *Certification of Disclosure in Companies' Quarterly and Annual Reports.* Release Nos. 33-8124, 34-46427. Washington, D.C.: SEC.

U.S. House of Representatives. 2002. The Sarbanes-Oxley Act of 2002. Public Law 107-204 [H. R. 3763]. Washington, D.C.: Government Printing Office.

Footnotes: Footnotes are not used for documentation. Textual footnotes should be used only for extensions and useful excursions of information that if included in the body of the text might disrupt its continuity. Footnotes should be inserted using the "footnote" or "endnote" feature of the word processing software which will automatically number the footnotes throughout the manuscript with superscript Arabic numerals.

SUBMISSION OF MANUSCRIPTS

Authors should note the following guidelines for submitting manuscripts:

1. Manuscripts currently under consideration by another journal or publisher should not be submitted. The author must state upon submission that the work is not submitted or published elsewhere.
2. For manuscripts reporting on field surveys or experiments: If the additional documentation (e.g., questionnaire, case, interview schedule) is sent as a separate file, then all information that might identify the author(s) must be deleted from the instrument.
3. Through May 31, 2011, manuscripts should be submitted via email as Microsoft Word or PDF file to the Senior Editor, Steven Kachelmeier, at email address tar@mccombs.utexas.edu. Please submit separate files for (1) the manuscript's title page with identifying information (not forwarded to reviewers), (2) the manuscript with title page and all other identifying information removed, and (3) any necessary supplemental files, such as experimental instructions and/or response memoranda on invited revisions. The body or subject line of the email should also indicate the submission fee receipt code from AAA (see step 5 below).
4. Beginning June 1, 2011, please check online at https://aaahq.org/AAAforms/journals/tarsubmit.cfm for instructions on how to submit a new manuscript to the incoming editor, John Harry Evans III. Revisions of manuscripts originally submitted before June 1, 2011 should be sent via email to the incoming editor, John Harry Evans III, at tar@katz.pitt.edu.
5. A nonrefundable submission fee is required unless explicitly waived by the editor for invited revisions of previous submissions. The submission fee in U.S. funds is $200.00 for members and $400.00 for nonmembers of the AAA payable by credit card (VISA or MasterCard only). The payment form is available online at: https://aaahq.org/AAAforms/journals/tarsubmit.cfm. If you are unable to pay by credit card or have any questions, please contact the AAA Member Services Team at (941) 921-7747 or info@aaahq.org.
6. Revisions must be submitted within 12 months from the decision letter inviting a revision.

COMMENTS

The Accounting Review welcomes submissions of comments on previous *TAR* articles. Comments on articles previously published in *The Accounting Review* will generally be reviewed by two reviewers, usually including an author of the original article (to assist the editor in evaluating whether the submitted comment represents the prior article accurately) and an additional reviewer who is independent of the original article. If a comment is accepted for publication, the original author will generally be invited to reply. All other editorial requirements, as enumerated above, apply to proposed comments.

POLICY ON REPRODUCTION

POLICY ON DATA AVAILABILITY

The AAA's Executive Committee policy (originally adopted in 1989, and amended in 2009) is that the objective of the Association-wide journals (*The Accounting Review, Accounting Horizons, Issues in Accounting Education*) is to provide the widest possible dissemination of knowledge based on systematic scholarly inquiries into accounting as a field of professional research, and educational activity. To fulfill this objective, authors are encouraged to make their data available for use by others in extending or replicating results reported in their articles.

MAY 2011 PLACEMENT ADS

The deadline for free position ads to be included in this section is two months prior to the date of publication. *The Accounting Review* is published in January, March, May, July, September, and November. Position ads in the "Placement Ads" are free with the purchase of a job posting in the AAA Career Center. For more information on how to purchase a job posting in the Career Center and receive your free position ad in *The Accounting Review*, go to the AAA website at http://aaahq.org and click on "Career Center," or call our office at 941-921-7747.

301 THE UNIVERSITY OF GEORGIA, Department of Housing and Consumer Economics College of Family and Consumer Sciences is seeking candidates for an Endowed Professor in Family Financial Planning with an emphasis in estate planning, planned giving, or comprehensive financial planning. Nine-month academic appointment. An endowed professorship is associated with this position. Position begins August 1, 2011. Qualifications: Significant academic contributions to the field of financial planning. Demonstrated record of outstanding scholarly achievement in research productivity, external funding, and successful college teaching experience desired. Current academic rank should be that of Professor, or work should be commensurate with the rank of Professor. For a complete position description and more information, please visit our website at: http://www.hr.uga.edu/hace-prof.pdf. Application Procedure: To receive full consideration, applications should be submitted prior to April 3, 2011; however, the position will remain open until filled. Application materials should include a cover letter of interest, which outlines relevant experience, supports your fit for the position, as well as a current *curriculum vitae*. Email submissions with attachments are preferred. Please send to: executivesearch@uga.edu. Confidential requests for information should be directed to: Executive Michael Luthi, Search Consultant, Executive and Faculty Search Group, (706) 542-1837. The University of Georgia is an Equal Opportunity/Affirmative Action Institution.

302 THE UNIVERSITY OF ALABAMA, Culverhouse School of Accountancy invites applications for a tenure-track appointment at the Assistant Professor level beginning August 16, 2011. Candidates should have an earned doctorate (or expect to complete their doctorate by August 2011) and exhibit excellence in teaching and the capacity for research publishable in top-tier accounting journals. The faculty member can expect to teach undergraduate- and graduate-level accounting courses, although current needs are driven by increasing graduate enrollment. The faculty member will have the ability to work with doctoral students and engage in University, College, and Departmental service activities. We welcome applications in all areas of accounting research and teaching. Screening of candidates will begin immediately and will continue until the position is filled. Salary is competitive. Candidates selected for interview will be required to submit a disclosure and consent form authorizing a background investigation. Only online applications are accepted. Candidates must apply online at: https://facultyjobs.ua.edu. Required documents include: resume/*curriculum vitae*, dissertation proposal and/or example of scholarly work, and a listing of four references. The University of Alabama is an Equal Opportunity Educational Institution/Employer. Women and Minorities are encouraged to apply.

303 CENTRAL CONNECTICUT STATE UNIVERSITY, Accounting Department invites applications for a tenure-track Assistant/Associate Professor (#C11-058) beginning Fall 2011. Teaching interest in financial accounting and/or auditing is preferred. Candidates must have a commitment to excellence in teaching and scholarship. Qualifications: Earned doctorate in accounting or equivalent from an accredited school preferred, ABD with a 2011 expected completion date required. Professional certifications and/or experience are desirable. Teaching interest in financial accounting and/or auditing is preferred. Candidates must have a commitment to excellence in teaching and scholarship. To apply: Complete the Application Cover Page at: http://www.ccsu.edu/AffAction/. Send the cover page, along with a letter of application, your resume, and the names, addresses, and telephone numbers of three references via email to: Professor Lawrence Grasso, GrassoLa@ccsu.edu. Applications received by February 15, 2011 will receive preference of review. CCSU aggressively pursues a program of equal employment and educational opportunity and affirmative action. Members of all underrepresented groups, women, veterans, and persons with disabilities are invited and encouraged to apply.

304 UNIVERSITY OF SOUTH CAROLINA, School of Accounting is accepting applications for a tenure-track position beginning Fall 2011. Candidates should have primary teaching and research interests in international accounting, and must possess a doctorate or expect to complete it by August 2011. In addition to being committed to excellent teaching, candidates must be ambitious and motivated researchers who are committed to publishing in top-tier accounting journals. Salaries are competitive and commensurate with experience and achievements. All faculty searches are subject to the availability of funding. Submit *curriculum vitae*, working papers, and defended dissertation proposal with data to: Professor Al Leitch, Darla Moore School of Business, University of South Carolina, Columbia, SC 29208; or Email: leitch@moore.sc.edu. The University of South Carolina is an AA/EOE. Minorities and women are strongly encouraged to apply.

305 SAINT LOUIS UNIVERSITY, a Catholic, Jesuit institution dedicated to student learning, research, health care, and service, is seeking applicants for a tenure-track position in Accounting at the Senior Assistant/Associate/Full Professor level, beginning Fall 2011. Candidate must demonstrate a commitment to teaching excellence and quality research and publication. Service to the department and University are also required in addition to teaching and scholarship. Professional certification is desirable. Teaching interest is open and all areas will be considered except financial accounting. A Ph.D. in Accounting is required. Dissertation-stage candidates could be considered under exceptional circumstances. Compensation package is competitive. Interested candidates must apply online at http://jobs.slu.edu (Req. 20100775), and submit a cover letter and current *curriculum vitae*. All other correspondence regarding this position can be sent to: Dr. Ananth Seetharaman, Chair of Accounting, John Cook School of Business, Saint Louis University, 3674 Lindell Boulevard, St. Louis, MO 63108. Saint Louis University is an Affirmative Action/Equal Opportunity Employer, and encourages nominations of and application from women and minorities.

306 SABANCI UNIVERSITY (SU), School of Management invites applications for full-time/visiting positions in Accounting for the year 2011–2012. While pre-doctorate candidates are considered, the ideal candidate possesses a Ph.D. and a track record of excellence in research and teaching. SU, a private, innovative academic institution, offers excellent facilities and a light teaching load amid a modern campus located in Istanbul, Turkey. The University admits top-ranking faculty and students and offers undergraduate, M.B.A., Executive M.B.A. and Ph.D. programs in Operations Management, Organization Studies, and Finance. It provides faculty excellent support and competitive compensation, including housing. English is the medium of instruction. For further information, visit: http://www.sabanciuniv.edu/ybf/eng/. Applicants are requested to complete the online application form at: https://crm.sabanciuniv.edu/acadcv/application.php, using ACC2011 as the Code Reference, and to send their *cv* and the names of three references to: Ulf Nilsson, nilsson@sabanciuniv.edu or Mine Aksu, maksu@sabanciuniv.edu. We will be interviewing at the Annual Meeting.

307 QUINNIPIAC UNIVERSITY invites applications for a full-time, tenure-track accounting faculty position, rank open, to start in August 2011. Consideration will be given to individuals interested in a full-time visiting position for the 2011–12 academic year. Primary teaching responsibilities will be in the financial accounting area. All candidates must hold a doctorate in accounting from an AACSB-accredited school (ABD considered for the visiting position). Successful candidates must demonstrate a strong potential for or have an established record of excellent teaching, published research, and significant service. Qualified candidates should apply online at https://careers.quinnipiac.edu. Application materials should include a *curriculum vitae*, evidence of teaching success, and the names of three references. Other inquiries should be directed to: Kathleen Simione, Chair of the Accounting Department at: accountingsearch@quinnipiac.edu. Quinnipiac University has a strong commitment to the principles and practices of diversity throughout the University community and welcomes candidates who would enhance that diversity.

308 OREGON STATE UNIVERSITY invites applicants for two anticipated full-time (1.0 FTE), nine-month, tenure-track positions in Accounting at the rank of Assistant Professor to begin September 2011. Teaching responsibilities include graduate and undergraduate accounting courses. Scholarly activities are expected to result in peer-reviewed journal publications. Required qualifications include an appropriate earned doctorate from an AACSB International accredited university or equivalent and evidence of potential for scholarly publication and teaching effectiveness. Preferred qualifications include professional certification and a demonstrable commitment to promoting and enhancing diversity. Information about the position and the University's online application process can be accessed at: http://oregonstate.edu/jobs, Posting No. 006857. Applications, including (1) a letter of interest clearly indicating interest in the position, (2) *cv*, (3) contact information for three professional references, (4) recent teaching evaluations and summary information of courses taught, and (5) evidence of scholarly work should be submitted by 3/15/2011. OSU is an AA/EOE.

309 TOWSON UNIVERSITY, College of Business and Economics invites applications for the position of Chair of the Department of Accounting, effective Summer or Fall 2011. The College of Business and Economics is an AACSB-accredited institution with more than 3,200 undergraduate business majors. The Department of Accounting has 18 full-time faculty members, more than 600 undergraduate majors, and more than 60 graduate students in a joint master's program in accounting with the University of Baltimore. Additional information can be found at: http://wwwnew.towson.edu/cbe/. Applicants should possess a Ph.D./D.B.A. with a concentration in accounting or closely related field from an AACSB-accredited university, and must provide evidence of effective teaching and intellectual contributions. Experience with the AACSB-accreditation process and requirements are highly desirable. This is a renewable 12-month administrative appointment at the Clinical Associate Professor level for an expected initial period of five years, with a reduced teaching load. Compensation is competitive. Review of applications will begin immediately. Interested candidates should submit a letter of application, current *curriculum vitae*, recent teaching evaluations (if applicable), and the names of three references to: Dr. Andrew Schiff, Search Committee Chair, Towson University, Department of Accounting, Stephens Hall, Room 102, 8000 York Road, Towson, MD 21252-0001, or Email: aschiff@towson.edu. Towson University is an Equal Opportunity/Affirmative Action Employer and has a strong institutional commitment to diversity. Women, minorities, persons with disabilities, and veterans are encouraged to apply.

310 WEST VIRGINIA UNIVERSITY, Department of Accounting and MIS in the College of Business and Economics invites applications for a tenure-track appointment at the Assistant or Associate Professor level beginning in August 2011. Our primary need is in Accounting Information Systems, but we welcome applications in all areas of accounting research and teaching. We anticipate filling at least one position. West Virginia University is a land-grant institution located in Morgantown, West Virginia, close to major metropolitan centers in the northeast. The Department of Accounting and MIS is separately accredited by the AACSB and offers undergraduate and graduate degrees, including a planned doctoral program. Applicants should have an earned doctorate in accounting (or equivalent) from an AACSB-accredited institution and exhibit effectiveness in teaching and the capacity for research publishable in respected scholarly outlets. CPA or other professional certification preferred. Interested candidates should submit a current *vita*, a research paper suitable for presentation at a workshop, the names and contact information of three references, and evidence of teaching ability to: Dr. Barbara Apostolou, Chair of the Search Committee, via email at: Barbara.Apostolou@mail.wvu.edu. Screening of candidates will begin immediately and continue until the position is filled. West Virginia University is an Equal Employment Opportunity/ Affirmative Action Institution. The university does not discriminate on the basis of age, color, disability, national origin, race, religion, sexual orientation, or veteran status. West Virginia University is the recipient of an NSF ADVANCE award for gender equity. Minorities, persons with disabilities, females, and other protected class members are encouraged to apply. This offer of employment is conditioned upon approval by all the appropriate governmental authorities.

311 GEORGIA COLLEGE & STATE UNIVERSITY, J. Whitney Bunting College of Business is accepting applications for an academic-year tenure-track position in accounting beginning August 2011. The successful candidate will be able to teach introductory, upper-level, and graduate accounting courses. Preference will be given to candidates with a record of successful teaching in the area of financial accounting. Qualified candidates will have completed, or be near completion, of a Ph.D. or D.B.A. degree in Accounting from a business school with AACSB Accreditation. Professional experience and certification are considered very positive complements to academic qualifications. Excellence in teaching, scholarly activity, and university/community service are requirements for promotion and tenure. To be considered, applicants must apply electronically at: http://www.gcsujobs.com

312 ROBERT MORRIS UNIVERSITY (RMU), founded in 1921, has offered high-quality academic programs in business to students from the Pittsburgh region. RMU's School of Business, accredited by AACSB International, seeks qualified candidates with teaching interests in financial/managerial/ cost accounting or accounting information systems. Qualified candidates should have an earned doctorate in Accounting from an AACSB-accredited school or be ABD near completion. Candidates must demonstrate excellence in teaching and research in their disciplines. Candidates should submit a letter of interest, current *vita*, and three academic and/or professional references to: Business@rmu.edu, c/o Dr. Jacobs, Dean, or mail to: Robert Morris University, 6001 University Boulevard, Moon Township, PA 15108, c/o Dr. Jacobs. Visit http://www.rmu.jobs for position descriptions. Anticipated start date is Fall 2011 and dependent upon budgetary approval. RMU is committed to increasing diversity in its community and actively pursues individuals from all backgrounds. EOE

313 PURDUE UNIVERSITY, Krannert Graduate School of Management invites applicants for one or more Continuous Term Lecturer appointment(s) in accounting beginning Fall 2011. Candidates will be expected to teach both Introduction to Tax and Advanced Taxation and be capable of teaching both financial and managerial accounting, primarily at the undergraduate level. A full-time teaching load is typically 24 credit hours per year. A successful candidate will have a demonstrated record of excellence in teaching preferably at comparable schools. A minimum of a master's degree with a professional certification in accounting is required; a Ph.D. with a concentration in accounting or related area is strongly preferred. Applications will be reviewed beginning February 15, 2011 and should include a *vitae*, evidence of teaching performance, and three letters of reference. Applications should be sent to: Professor Susan Watts, Krannert Graduate School of Management, Purdue University, 403 W. State Street, West Lafayette, IN 47907-2056. Purdue University is an Equal Opportunity/Equal Access/Affirmative Action Employer fully committed to achieving a diverse workforce.

314 UNIVERSITY OF WISCONSIN–STEVENS POINT is seeking candidates for an Assistant Professor of Accounting. The successful candidate will teach introductory and advanced courses in Accounting, with preference given to an interest in Auditing. Prefer Ph.D./D.B.A. in accounting from an AACSB-accredited institution. ABD will be considered. Ph.D. required for tenure. Send a letter of application including primary areas of teaching interest, resume or *vita*, copies of transcripts, and a list of three references (including phone numbers) to: Dr. Gary Mullins, Head School of Business and Economics, University of Wisconsin–Stevens Point, Stevens Point, WI 54481-3897; Email: gmullins@uwsp.edu. Screening will begin February 22, 2011. Position open until filled. Please see full position announcement at: http://www.UWSP.edu.

315 UNIVERSITY OF WASHINGTON BOTHELL, Business Program invites applications for a tenured opening in Accounting at the Professor or Associate Professor level. The appointment will become effective in the Fall 2011. Applicants must have a Ph.D. in Accounting and a commitment to excellence in scholarship and teaching. University of Washington faculty members engage in research, teaching, and service. Candidates should expect to teach at all levels of the undergraduate and graduate curriculum. They must be committed to working with diverse student and community populations. This position is contingent upon available funding. The successful candidate will have an established record of publications in high-quality academic journals and demonstrated excellence in teaching. In addition, he or she will play an important role in the continued development of the recently launched Business Program Accounting Option and new initiatives in the accounting area. This will include program and curriculum development, scheduling, mentoring junior faculty, and hiring part-time faculty. To apply: Please send a letter describing your qualifications and interest in the position as well as your *curriculum vitae* to: Dr. Steve Holland Chair, Accounting Search Committee, Business Program, Box 358533, University of Washington Bothell, 18115 Campus Way NE, Bothell, WA 98011-8246; Phone: (425) 352-5232; Fax: (425) 352-5277; Email: sholland@uwb.edu. For additional information, please see our website at: http://www.uwb.edu/bus/.

316 TEXAS CHRISTIAN UNIVERSITY, M. J. Neeley School of Business, Department of Accounting invites applications for a second Assistant Professor position to begin in August 2011. Candidates should hold a doctoral degree or expect to complete their degree prior to the start date. We seek candidates who are both excellent teachers and excellent researchers, consistent with TCU's Teacher-Scholar model. Applications should include a cover letter, *curriculum vitae*, and letters from at least three references. Applications may be submitted in either hardcopy or electronically, although electronic copies are preferred. Please send applications or nominations to: Ray Pfeiffer, Department Chair & Professor of Accounting, Texas Christian University, M. J. Neeley School of Business, TCU Box 298530, Fort Worth, TX 76129: Phone: (817) 257-6435: Email: r.pfeiffer@tcu.edu. Review of applications will continue until the position is filled. EEO/AA.

317 MISSISSIPPI STATE UNIVERSITY–MERIDIAN, Division of Business, is accepting applications for the position of Assistant Professor of Accounting. Teaching emphasis will be in managerial/ financial accounting. This is a tenure-track position beginning August 16, 2011. Minimum qualifications: Doctorate in Accounting from an AACSB-accredited institution or Doctorate in Business Administration with an emphasis in Accounting from an AACSB-accredited institution with demonstrated excellence in the field of accountancy. Preferred qualifications: Professional certification and practical experience in accounting. Responsibilities: Teach a variety of undergraduate and graduate courses in accounting, publishing high-quality research, and advising and mentoring students. Review of applications will begin immediately and continue until position is filled. Interested applicants must complete the Personal Data Information Form online at: http://www.jobs.msstate.edu, and send: (1) letter of application; (2) *vitae*; (3) official transcripts; (4) three letters of recommendation; and (5) the name, address, and telephone number of three additional references to: Dr. Kevin Ennis, Search Committee Chair, Division of Business, MSU–Meridian, 1000 Highway 19 North, Meridian, MS 39307. MSU is an AA/EOE.

THE ACCOUNTING REVIEW
ISSN 0001-4826

PERIODICALS POSTAGE PAID
SARASOTA FLORIDA
AND ADDITIONAL MAILING OFFICES

POSTMASTER
Send Address Changes To:

AMERICAN ACCOUNTING ASSOCIATION
5717 BESSIE DRIVE
SARASOTA, FL 34233